

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:05 ; Search time 8.84706 Seconds

276.623 Million cell updates/sec

Title: US-10-004-381-25

Sequence: 1 MDEKTTGMRGSHVVEGLAE.....LEHHPQGRREPMMGGCKLG 47

Scoring table: BLOSUM62

Gapor 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	60.5	23.4	177	1	HSLV_RHLLO	O98c8 rhizobium I
2	59.5	23.1	342	1	YDEY_ECOLI	P76762 escherichia d
3	57.5	22.3	1065	1	KDGI_HUMAN	O75912 homo sapiens
4	57	22.1	474	1	TRPC_COGL	P06560 cornuabact
5	56.5	21.9	215	1	GPME_SALT1	O820C4 salmonella
6	56.5	21.9	215	1	GPME_SALT1	O82J08 salmonella
7	56.5	21.9	404	1	KVB3_HUMAN	O43448 homo sapiens
8	56.5	21.9	415	1	BFS2_BOVIN	O28177 bos taurus
9	56.5	21.9	1560	1	SMCX_HUMAN	P41229 homo sapiens
10	56	21.7	261	1	YEVE_YEAST	P40078 saccharomyco
11	56	21.7	299	1	V152_STRCO	O92513 streptomyces
12	55.5	21.5	188	1	HSLV_CACRC	O9a239 caulobacter
13	55.5	21.5	288	1	AOX4_AERPE	O9y4d4 aeropyrum f
14	55.5	21.5	465	1	FXD3_MOUSE	O61060 mus musculu
15	54	20.9	376	1	ALC_STRCO	O93KJ4 streptomyces
16	53.5	20.7	352	1	GCP_TREPA	O83666 treponema p
17	53.5	20.7	404	1	KVB1_RAT	O63494 rattus norv
18	53	20.5	348	1	PLSX_SYNY3	P73950 synchocyst
19	53	20.5	358	1	AROB_AERPE	O9yej9 aeropyrum f
20	53	20.5	381	1	PTX3_MOUSE	P48759 mus musculu
21	53	20.5	483	1	MURE_CHICK	O823M2 chlamydophi
22	53	20.5	528	1	WR42_ARATH	O9xec3 arabidopsis
23	53	20.5	589	1	IFEB_ASGSU	P23731 ascaris suu
24	52.5	20.3	184	1	HSLV_BRUME	O8ye1 bruceella me
25	52.5	20.3	184	1	HSLV_BRUSU	O8y11 bruceella su
26	52.5	20.3	344	1	PLSX_SYNEL	O8dk15 synchococcc
27	52.5	20.3	461	1	SLX10_CHICK	O9w757 gallus galli
28	52.5	20.3	591	1	VATA_STRP3	O8k8r1 streptococcc
29	52	20.2	113	1	KRE_BACSU	P23789 bacillus su
30	52	20.2	270	1	BDNF_CYPCA	O90322 cyprinus ca
31	52	20.2	356	1	GBR2_CAEEL	O20616 caenorhabdi
32	52	20.2	378	1	HMAN_STROM	P02833 drosophila
33	52	20.2	449	1	AROF_STROE	P80574 streptomyces

34	52	20.2	499	1	A1NX_HUMAN	Q16332	homo sapiens
35	52	20.2	504	1	A1NX_MOUSE	P46660	mus musculus
36	52	20.2	505	1	A1NX_RAT	P23565	rattus norvegicus
37	52	20.2	2337	1	TOR2_SCHRO	Q93782	schizosaccharomyces pombe
38	51.5	20.0	181	1	COB1_ECOLI	P46886	escherichia coli
39	51.5	20.0	273	1	CG80_BACU	P80878	bacillus subtilis
40	51.5	20.0	323	1	CG36_HUMAN	O93344	homo sapiens
41	51.5	20.0	392	1	PRD1_STROR	P52321	streptococcus pneumoniae
42	51.5	20.0	401	1	KVBI_MOUSE	P63277	mus musculus
43	51.5	20.0	408	1	KVBI_MOUSE	Q28528	mus musculus
44	51.5	20.0	419	1	KVBI_HUMAN	Q14722	homo sapiens
45	51.5	20.0	419	1	KVBI_RABIT	O95t31	oryctolagus cuniculus

ALIGNMENTS

ID	HSLV_RHILLO	STANDARD	PRT	177 AA.
AC	098CT8			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	ATP-dependent protease hslV (EC 3.4.25.-).			
DE	HSLV OR MLL5007.			
OC	Rhizobium lotfi (Mesorhizobium lotfi).			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Phyllobacteriaceae; Mesorhizobium.			
OX	NCBI TaxID=381;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MAFF303099;			
RC	MEDLINE=21082930; PubMed=11214968;			
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,			
RA	Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,			
RA	Mochoizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,			
RA	Takeuchi C., Yamada M., Tabata S.;			
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium			
RT	Mesorhizobium lotfi";			
RL	DNA Res. 7:331-338(2000).			
CC	-1- FUNCTION: Protease subunit of a proteasome-like degradation			
CC	complex (By similarity).			
CC	-1- SUBUNIT: A double ring-shaped homohexamer of hslV is capped on			
CC	each side by a ring-shaped hslU homohexamer (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	DR EMBL; AP003005; BAB5153.1; -			
CC	DR MEROPS; T01.006; -			
CC	DR HAMAP; MF_00248; -; 1.			
DR	InterPro; IPR001353; Peptidase_T1.			
DR	Pfam; PF00227; proteasome; 1.			
DR	HydroLase; Protease; Threonine protease; Complete proteome.			
FT	ACT_SITE 4			
FT	ACT_SITE 4			
SO	SEQUENCE 177 AA; 18986 MW; 62F3D3DC6ACB5F05 CRC64;			
Query Match	23.4%; Score 60.5; DB 1; Length 177;			
Best Local Similarity	41.9%; Pred. No. 1.7;			
Matches	13; Conservative 6; Mismatches 5; Indels 7; Gaps 1;			
OY	9 RGHHVEGLAGE-----LEQLRLLEHHH 32			
db	42 GGNVINGAGATADATFLRLERLAKKEOYP 72			

DR EMBL; AF219920; AAF43006.1; JOINED.
 DR EMBL; AF219921; AAF43006.1; JOINED.
 DR EMBL; AF219922; AAF43006.1; JOINED.
 DR EMBL; AF219923; AAF43006.1; JOINED.
 DR EMBL; AF219924; AAF43006.1; JOINED.
 DR EMBL; AF219925; AAF43006.1; JOINED.
 DR EMBL; AF219926; AAF43006.1; JOINED.
 DR EMBL; AF219927; AAF43006.1; JOINED.
 DR EMBL; AF219928; AAF43006.1; JOINED.
 DR EMBL; AF219929; AAF43006.1; JOINED.
 DR EMBL; AF219930; AAF43006.1; JOINED.
 DR EMBL; AF219931; AAF43006.1; JOINED.
 DR EMBL; AF219932; AAF43006.1; JOINED.
 DR EMBL; AF219933; AAF43006.1; JOINED.
 DR EMBL; AF219934; AAF43006.1; JOINED.
 DR EMBL; AF219935; AAF43006.1; JOINED.
 DR EMBL; AF219936; AAF43006.1; JOINED.
 DR EMBL; AF219937; AAF43006.1; JOINED.
 DR EMBL; AF219938; AAF43006.1; JOINED.
 DR GeneW; HGNC:2855; DGKI.
 DR MIM; 604072; .
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0004143; F:diacylglycerol kinase activity; TAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000756; DAGKa.
 DR InterPro; IPR001206; DAGKc.
 DR Pfam; PF00023; ank; 2.
 DR Pfam; PF00609; DAGKa; 1.
 DR Pfam; PF00781; DAGKc; 1.
 DR ProDom; PD002939; DAGKa; 1.
 DR ProDom; PD005043; DAGKc; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00109; C1; 2.
 DR SMART; SM00045; DAGKa; 1.
 DR SMART; SM00046; DAGKc; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 2.
 DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PSS0479; DAG_PE_BIND_DOM_1; FALSE_NEG.
 DR PROSITE; PSS0081; DAG_PE_BIND_DOM_2; FALSE_NEG.
 DR Transfaser; Kinase; ANK repeat; Repeat; Nuclear protein;
 KM Multigene family; Polymorphism.
 FT DOMAIN 178 232
 FT DOMAIN 251 309 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 374 500 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 526 683 CATALYTIC-A (POTENTIAL).
 FT REPEAT 958 990 CATALYTIC-B (POTENTIAL).
 FT REPEAT 997 1026 ANK 1.
 FT DOMAIN 20 31 ANK 2.
 FT DOMAIN 69 74 POLY-SER.
 FT DOMAIN 95 102 POLY-ALA.
 FT VARIANT 153 153 L -> F.
 FT VARIANT 153 153 /PFI5VAR 010190.
 SQ CONFLICT 160 160 A -> P (IN REF. 2).
 FT SEQUENCE 1065 AA; 116996 MW; B84971AA630A799 CRC64;

Query Match 22.3%; Score 57.5; DB 1; Length 1065;
 Best Local Similarity 25.9%; Pred. No. 28;
 Matches 14; Conservative 9; Mismatches 20; Indels 11; Gaps 1;

DB 4 KTTGWRGHH-----VVEGLAGELEQLARLEHHHGGGRRMMGGCKL 46
 469 KTLNMGGGYDEPVSKITLCOVEDGTAVQLDKRNLIHVERNDLPPEBLDGVCKL 522

RESULT 4
 TRPC CORGL STANDARD; PRT; 474 AA.
 AC P06560;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Trypctophan biosynthesis protein trpCF [includes: indole-3-glycerol
 phosphate synthase (EC 4.1.1.48) (IGPS); N-(5'-phospho-
 ribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI)].
 GN TRPC OR CGJ3033.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87117512; PubMed=3808947;
 RA Matsui K., Sano K., Ohtsubo E.;
 RT "Complete nucleotide and deduced amino acid sequences of the
 Brevibacterium lactofermentum tryptophan operon.";
 RL Nucleic Acids Res. 14:10113-10114(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
 CC - FUNCTION: Bifunctional enzyme that catalyzes two sequential steps
 of tryptophan biosynthetic pathway. The first reaction is
 catalyzed by the isomerase, coded by the trpF domain; the second
 reaction is catalyzed by the synthase, coded by the trpC domain.
 CC - CATALYTIC ACTIVITY: N-(5-phospho-beta-D-riboseyl)-anthranilate = 1-
 (2-carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate
 CC - CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribose 5-
 phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
 CC - PATHWAY: Tryptophan biosynthesis; third step.
 CC - PATHWAY: Tryptophan biosynthesis; fourth step.
 CC - SUBUNIT: Monomer.
 CC - SIMILARITY: In the N-terminal section; belongs to the trpC family.
 CC - SIMILARITY: In the C-terminal section; belongs to the trpF family.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04960; CAA28626.1; .
 DR EMBL; AP005283; BAC00427.1; .
 DR F1R; E24723; E24723.
 DR HSP; P00909; 1P11.
 DR HAMAP; MF_00134; fused; 1.
 DR HAMAP; MF_00135; fused; 1.
 DR InterPro; IPR001468; IGPS.
 DR InterPro; IPR001240; PRAI.
 DR Pfam; PF00218; IGPS; 1.
 DR Pfam; PF00697; PRAI; 1.
 DR ProDom; PD006151; IGPS; 1.
 DR PROSITE; PSS0614; IGPS; 1.
 KM Tryptophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
 KM Decarboxylase; Complete proteome.
 FT DOMAIN 1 262
 FT DOMAIN 263 474
 FT CONFLICT 88 89
 FT CONFLICT 110 110 A -> G (IN REF. 1).
 FT CONFLICT 130 131 HA -> RP (IN REF. 1).
 FT CONFLICT 153 153 A -> D (IN REF. 1).
 FT CONFLICT 302 302 L -> S (IN REF. 1).
 FT CONFLICT 343 343 D -> G (IN REF. 1).
 FT CONFLICT 381 383 MISSING (IN REF. 1).
 FT CONFLICT 454 474 AGAKDAGALKTKFATISTFHY -> GWGRCRAAENFRDH
 LHIPLKY (IN REF. 1).
 SQ SEQUENCE 474 AA; 50477 MW; C347C7016BB97F9A CRC64;

Query Match 22.1%; Score 57; DB 1; Length 474;
 Best Local Similarity 55.6%; Pred. No. 14;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 13 YVEGLAGLEQLRAREH 30
:|||||:|||||:
Db 13 IVEGRGHLEIRARIAH 30

RESULT 5

GPMB_SALTY STANDARD; PRT; 215 AA.
AC 08Z0T4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable phosphoglycerate mutase gpmb (EC 5.4.2.1)
DE (Phosphoglyceromutase) (PGAM).
GN GPMB OR ST14932 OR T4624.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=601;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21534947; PubMed=11677608;
RA Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Parkhill J., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagals K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhimurium CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burdand V., Kodyranni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomes of Salmonella enterica serovar Typh strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
= 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC -1- PATHWAY: Glycolysis.
CC -1- SIMILARITY: Belongs to the phosphoglycerate mutase family. Gpmb
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627284; CAD03416.1; -;
DR EMBL; AE016849; AA072055.1; -;
DR HAMAP; MF_01040; -; 1.
DR InterPro; IPR001345; PG/BPGM_mutase.
DR Pfam; PF00300; PGAM; 1.
DR PROSITE; PS00175; PG_MUTASE; 1.
KW Isomerase; Glycolysis; Complete proteome.
FT ACT_SITE 9
FT ACT_SITE 9
FT ACT_SITE 9
FT ACT_SITE 58
FT ACT_SITE 58
FT ACT_SITE 150
FT ACT_SITE 150
FT ACT_SITE 215 AA; 23958 MW; BC47D35F40F819 CRC64;
SQ SEQUENCE 215 AA; 23958 MW; BC47D35F40F819 CRC64;

Query Match 21.9%; Score 56.5; DB 1; Length 215;

Best Local Similarity 26.7%; Pred. No. 6.8; Mismatches 23; Indels 13; Gaps 1;

QY 1 MDEKTTGR-----GGHVEGLAGLEQLRAREHHPGGRBPMMSGGCKLG 47
:|||||:|||||:|||||:|||||:
Db 96 LTBEEBGRQQLVNGTQDGRIPGSEMOELSDRYAALASCLERPGSRPLVSHGIALG 155

RESULT 6

GPMB_SALTY STANDARD; PRT; 215 AA.
AC 08Z0T8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable phosphoglycerate mutase gpmb (EC 5.4.2.1)
DE (Phosphoglyceromutase) (PGAM).
GN GPMB OR STW4585.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX McClelland M., Sanderson K.E., Speeth J., Clifton S.W., Latreille P.,
RA McClelland M., Sanderson K.E., Speeth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
= 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC -1- PATHWAY: Glycolysis.
CC -1- SIMILARITY: Belongs to the phosphoglycerate mutase family. Gpmb
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE008915; AAL23400.1; -;
DR STyGene; SG27272; gpmb.
DR HAMAP; MF_01040; -; 1.
DR InterPro; IPR001345; PG/BPGM_mutase.
DR Pfam; PF00300; PGAM; 1.
DR PROSITE; PS00175; PG_MUTASE; 1.
KW Isomerase; Glycolysis; Complete proteome.
FT ACT_SITE 9
FT ACT_SITE 9
FT ACT_SITE 9
FT ACT_SITE 58
FT ACT_SITE 58
FT ACT_SITE 150
FT ACT_SITE 150
FT ACT_SITE 215 AA; 23868 MW; ADFD335F50FF809 CRC64;
SQ SEQUENCE 215 AA; 23868 MW; ADFD335F50FF809 CRC64;
Query Match 21.9%; Score 56.5; DB 1; Length 215;
Best Local Similarity 26.7%; Pred. No. 6.8; Mismatches 23; Indels 13; Gaps 1;
Matches 16; Conservative 8; Mismatches 23; Indels 13; Gaps 1;
QY 1 MDEKTTGR-----GGHVEGLAGLEQLRAREHHPGGRBPMMSGGCKLG 47
:|||||:|||||:|||||:|||||:
Db 96 LTBEEBGRQQLVNGTQDGRIPGSEMOELSDRYAALASCLERPGSRPLVSHGIALG 155
RESULT 7

KVB3_HUMAN STANDARD; PRT; 404 AA.
 ID KVB3_HUMAN
 AC 043448;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3).
 GN KCNA3 OR KCNA3B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RX MEDLINE=9074289; PubMed=9857044;
 RA Leichter T., Baehrting R., Isbrandt D., Pongs O.;
 RT "Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel A-type potassium channel.";
 RL J. Biol. Chem. 273:35095-35101(1998).
 CC -1- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.5.
 CC -1- SUBUNIT: Forms heteromultimeric complex with alpha subunits.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- TISSUE SPECIFICITY: Brain-specific expression. Most prominent expression in cerebellum. Weaker signals detected in cortex, occipital lobe, frontal lobe and temporal lobe. Not detected in spinal cord, heart, lung, liver, kidney, pancreas, placenta and skeletal muscle.
 CC -1- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).
 CC -1- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF016411; AAB92499.1; -
 CC GenBank: AF016411; AAB92499.1; -
 CC MIM: 604111; -
 DR GO: GO:0015459; F:potassium channel regulator activity; TAS.
 DR GO: GO:0006813; P:potassium ion transport; TAS.
 DR InterPro: IPR001395; Aldo/ket_red.
 DR InterPro: IPR005402; KCNA3_channel.
 DR InterPro: IPR005399; KCNA3_channel.
 DR InterPro: IPR005983; KCNA3_core.
 DR Pfam: PF00248; aldo_ket_red; 1.
 DR PRINTS: PR01577; KCNA3CHANNEL.
 DR PRINTS: PR01580; KCNA3CHANNEL.
 DR PRODOM: PD000248; Aldo/ket_red; 2.
 DR TIGRAME: TIGR01293; Kv_beta; 1.
 DR Ionic channel; Ion transport; Potassium transport;
 KW Voltage-gated channel.
 KW Voltage-gated channel.
 SQ SEQUENCE 404 AA; 43530 MW; 08265CC07929A1BA CRC64;

Query Match 21.9%; Score 56.5; DB 1; Length 404;
 Best Local Similarity 31.1%; Pred. No. 13;
 Matches 14; Conservative 5; Mismatches 3; Indels 23; Gaps 2;

4 KITGWR-----GG-----HYVEGLAGLEQLR 25
 145 KSGMGRSSVYITTKIFWGGQATBRLSKRKHIIELGRLSLRLQ 189

RESULT 8
 BFS2_BOVIN

ID BFS2_BOVIN STANDARD; PRT; 415 AA.
 AC 028177;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phaknin (beaded filament structural protein 2) (Lens fiber cell beaded filament protein CP 49) (CP49) (49 kDa cytoskeletal protein) (CP 47) (CP47).
 GN BFS2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine; Bos.
 OC NCBI_TaxID=9913;
 RX MEDLINE=94075412; PubMed=7504675;
 RA Merdes A., Gounari F., Georgatos S.D.;
 RT "The 47-kD lens-specific protein phaknin is a tailless intermediate filament protein and an assembly partner of filensin.";
 RL J. Cell Biol. 123:1507-1516(1993).
 CC [2]
 CC REVISIONS.
 CC Merdes A.;
 CC Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: ASSOCIATES WITH BFS1.
 CC -1- SUBCELLULAR LOCATION: Membrane- and cytoskeleton-associated.
 CC -1- TISSUE SPECIFICITY: Lens.
 CC -1- SIMILARITY: Belongs to the intermediate filament family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: X75160; CAA53003.1; -
 CC InterPro: IPR001664; IF;
 DR InterPro: IPR002957; Keratin_1.
 DR Pfam: PF00038; filament; 1.
 DR PRINTS: PR01248; TYPEKERATIN.
 DR PROSITE: PS00226; IF; FALSE_NEG.
 DR Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;
 KW Eye lens protein.
 KW Eye lens protein.
 FT DOMAIN 1 114 HEAD.
 FT DOMAIN 115 395 ROD.
 FT DOMAIN 396 415 TAIL.
 FT DOMAIN 115 144 COILED COIL (POTENTIAL).
 FT DOMAIN 199 248 COILED COIL (POTENTIAL).
 FT DOMAIN 295 395 COILED COIL (POTENTIAL).
 SQ SEQUENCE 415 AA; 45949 MW; F3996852E6961B6 CRC64;

Query Match 21.9%; Score 56.5; DB 1; Length 415;
 Best Local Similarity 41.7%; Pred. No. 14;
 Matches 15; Conservative 5; Mismatches 15; Indels 1; Gaps 1;

11 GHVEGLAGLEQLRLARLEHHHPQGRPEPMGCGCKL 46
 364 GAVSRLEELAELEMRABEQQLA-REHLISHKQDL 398

RESULT 9
 SMCX_HUMAN STANDARD; PRT; 1560 AA.
 ID SMCX_HUMAN
 AC P41229;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Smcx protein (Xe169 protein).
 GN SMCX OR XE169.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxId=9605;
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=94214434; PubMed=8162017;
 RA Wu J., Ellison J., Salido E., Yen P., Mohandas T., Shapiro L.J.;
 RT "Isolation and characterization of XE169, a novel human gene that
 escapes X-inactivation.";
 RL Hum. Mol. Genet. 3:153-160 (1994).
 RN [2]
 RP SEQUENCE OF 280-344 FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=95038739; PubMed=7951230;
 RA Aguilnik A.I., Mitchell M.J., Mattei M.-G., Bersani G., Avner P.A.,
 RT "A novel X gene with a widely transcribed Y-linked homologue escapes
 X-inactivation in mouse and human.";
 RL Hum. Mol. Genet. 3:879-884 (1994).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P41229-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P41229-2; Sequence=VSP_000315;
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined. Highest
 CC levels found in skeletal muscle.
 CC -1- MISCELLANEOUS: Escapes X-inactivation.
 CC -1- SIMILARITY: Contains 1 ARID domain.
 CC -1- SIMILARITY: Contains 1 JMJN domain.
 CC -1- SIMILARITY: Contains 1 JMJN domain.
 CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L25270; AAA61302.1; -
 CC EMBL; 229650; CAA82758.1; -
 CC PIR; I54361; I54361.
 CC Gene; HGNC:11114; SMCX.
 DR MIM; 314690; -
 DR InterPro; IPR001606; ARID.
 DR InterPro; IPR003347; TF_JmjC.
 DR InterPro; IPR003349; TF_JmjN.
 DR InterPro; IPR004198; Znf_C5HC2.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF01388; ARID; 1.
 DR Pfam; PF02373; JmjC; 1.
 DR Pfam; PF02375; JmjN; 1.
 DR Pfam; PF00628; PHD; 2.
 DR Pfam; PF02928; zf-C5HC2; 1.
 DR SMART; SM00501; BRIGH1; 1.
 DR SMART; SM00558; JmjC; 1.
 DR SMART; SM00545; JmjN; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; ZF_PHD_2; 1.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 DR Zinc-finger; Repeat; Alternative splicing.
 KM
 FT DOMAIN 13 59 JMJN.
 FT DOMAIN 76 184 ARID.
 FT ZN_FING 326 372 PHD-TYPE 1.
 FT DOMAIN 501 617 JMJC.
 FT ZN_FING 1187 1248 PHD-TYPE 2.
 FT VARSPIC 1370 1372 Missing (in isoform 2).
 FT
 FT CONFLICT 342 342 /FTId=VSP_000315.
 FT C -> Y (IN REF. 2).

SQ SEQUENCE 1560 AA; 175804 MW; 78913F010DBCE1 CRC64;
 Query Match 21.9%; Score 56.5; DB 1; Length 1560;
 Best Local Similarity 28.9%; Pred. No. 57;
 Matches 13; Conservative 7; Mismatches 18; Indels 7; Gaps 1;
 QY 1 MDEKTYGMRG-----GHVVEGLAGELEQRLARLEHHPPGCGRRP 38
 DB 1281 LTERAISWGGRARQALASSEDYTAGLRLAEKRLQALAPRRERPP 1325
 RESULT 10
 YE6 YEAST
 ID YE6 YEAST STANDARD; PRT; 261 AA.
 AC P40078;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 29.7 kDa protein in RSP5-LCP5 intergenic region.
 GN YER126C OR SYCP-ORF47.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OK NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313364; PubMed=9169868;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Yelton M.A., Allen E.,
 RA Arayo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
 RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartwell G.,
 RA Huntke-Smith S., Hyman R., Kayser A., Komp C., Laibkari D., Lew H.,
 RA Lin D., Moseveld D., Nakahara K., Namath A., Norgren R., Oefner P.,
 RA Oh C., Petzel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
 RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
 RL Nature 387:78-81 (1997).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U18916; AAC03224.1; -
 CC PIR; S43218; S43218.
 DR GerMOnline; 139205; -
 DR SGD; S0000928; KRS32.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0042273; P:ribosomal large subunit biogenesis; IMP.
 DR InterPro; IPR001047; Ribosomal_S85.
 DR Pfam; PF01201; Ribosomal_S85; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 261 AA; 29722 MW; 545CB84250BCB0C9 CRC64;
 Query Match 21.7%; Score 56; DB 1; Length 261;
 Best Local Similarity 26.7%; Pred. No. 9.7;
 Matches 12; Conservative 10; Mismatches 19; Indels 4; Gaps 1;
 QY 3 EKTGWRGCHVVEGLAGELEQRLARLEHHPPGCGRRP 43
 DB 40 OKLTGWRGCHVVEGLAGELEQRLARLEHHPPGCGRRP 84
 RESULT 11
 Y52 STRCO
 ID Y52 STRCO STANDARD; PRT; 299 AA.
 AC Q9Z513;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Hypothetical UPF0042 protein SC01952.
GN SC01952 OR SCC54.12C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2).
RA Rydning N.J., Ainsa J.A., Hartley N., Bruton C.J., Chater K.F.;
RT "The whaA sporulation regulatory gene of Streptomyces coelicolor
A3(2)."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.T.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierzbicki A., Woodward J., Barrrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC -1 SIMILARITY: Belongs to the UPF0042 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF106003; AAD42167.1; ALT_INIT.
CC EMBL: AL939110; CAB38142.1; ALT_INIT.
CC PIR: T36025; T36025.
DR HAMAP: MF_00636; -; 1.
DR InterPro: IPR005337; UPF0042.
DR Pfam: PF03668; ATP_bind2.1.
KM Hypothetical protein; ATP-binding, Complete proteome.
FT NP_BIND 23 30 ATP (POTENTIAL).
SQ SEQUENCE 299 AA; 32821 MW; E69CE8455948B7CD CRC64;

Query Match 21.7%; Score 56; DB 1; Length 299;
Best Local Similarity 27.0%; Pred. No. 11;
Matches 17; Conservative 9; Mismatches 11; Indels 26; Gaps 2;

QY 11 GHHVEGLAGE-----QLRLRLEHHPGGRP-----NMSSGC 44
DB 127 GRVDSIAERELRLRGADLVITSSLVNHELRKADQAGQGEFLRATVWSFGR 186
QY 45 KLG 47
DB 187 KYG 189

RESULT 12
HSLV CAUCR STANDARD; PRT; 188 AA.
ID HSLV CAUCR
AC Q9A239;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent protease hslV (EC 3.4.25.-).
GN HSLV OR CC3727.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;

OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eissen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ueberback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1 FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1 SUBUNIT: A double ring-shaped homohexamer of hslV is capped on
CC each side by a ring-shaped hslU homohexamer (By similarity).
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1 SIMILARITY: Belongs to peptidase family T1b. hslV subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE006030; AAK25689.1; -.
CC PIR: B87711; B87711.
DR HSSP: P31059; 1B94.
DR MEROPS: T01.006; -.
DR TIGR: CC3727; -.
DR HAMAP: MF_00248; -; 1.
DR InterPro: IPR001353; Peptidase_T1.
DR Pfam: PF00227; proteasome_1.
KM Hydrolyase, Protease, Threonine protease, Complete proteome.
FT ACT_SITE 14 14 BY SIMILARITY.
FT ACT_SITE 14 14
SQ SEQUENCE 188 AA; 19600 MW; ACDDC1FCB810D61 CRC64;

Query Match 21.5%; Score 55.5; DB 1; Length 188;
Best Local Similarity 43.3%; Pred. No. 7.8;
Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;

QY 10 GGHVEGLAGE-----LEQLRLRLEHHP 32
DB 52 GSKVAGFAGATADAFILIERLEAKLEQYP 81

RESULT 13
AOK4 AERPE STANDARD; PRT; 288 AA.
ID AOK4 AERPE
AC Q9YDX4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme-copper oxidase subunit IV (EC 1.9.3.-).
GN AOKC OR APE0795.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE OF 186-288 FROM N.A.
RC STRAIN=K1;
RA Makagi T., Ishikawa R.;
RT "Heme-copper-oxidase";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;

```

RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosegi H.,
RA Hosoyama A., Fukui S., Nagai Y., Mshijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.",
RL DNA Ref. 6:83-101(1999).
CC
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC DR EMBL; AB020482; BAA86073.1; -.
CC DR EMBL; AB000060; BAA79773.1; -.
CC DR PIR; E72671; E72671.
CC KW Oxidoreductase; Transmembrane; Respiratory chain; Complete proteome.
CC FT TRANSMEM 205 225 POTENTIAL.
CC FT TRANSMEM 227 247 POTENTIAL.
CC FT TRANSMEM 260 280 POTENTIAL.
CC SQ SEQUENCE 288 AA; 31584 MW; 61B3DDDA610EC381 CRC64;

Query Match 21.5%; Score 55.5; DB 1; Length 288;
Best Local Similarity 28.8%; Pred. No. 12;
Matches 15; Conservative 4; Mismatches 18; Indels 15; Gaps 1;

QY 9 RCGHVVGLAGLEQLRLA-----RLHHPOGQREPMMSGCK 45
DB 108 RCGHVVGLCGDGTGTRTRHTGTSHTNISAPALQPGQSGATLRSQGR 159

RESULT 14
FXD3_MOUSE STANDARD; PRT; 465 AA.
AC Q61060;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forward box protein D3 (HNF3/TF transcription factor genes)
DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HNF-2).
GN FOXD3 OR HNF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBD databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Laboey P.A., Kaestner K.H.;
RT "The winged helix transcription factor Hth2 is expressed in neural
RT crest and spinal cord during mouse development.",
RL Submitted (MAY-1998) to the EMBL/GenBank/DBD databases.
CC -1- FUNCTION: Probable transcription factor.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC
CC DR EMBL; U41047; AAA87569.1; -.
CC DR EMBL; AF067421; AAC28352.1; -.
CC DR HSPB; Q63245; ZHRH.
CC DR TRANSFAC; T04166; -.
CC DR MGD; MGI:1347473; Foxd3.
CC DR InterPro; IPR001766; TF_Fork_head.
CC DR Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR ProDom; PD000425; TF_Fork_head; 1.
CC DR SMART; SM00339; FH; 1.
CC DR PROSITE; PS00657; FORK_HEAD_1; 1.
CC DR PROSITE; PS00658; FORK_HEAD_2; 1.
CC DR PROSITE; PS50039; FORK_HEAD_3; 1.
CC KW DNA-binding; Nuclear protein; Transcription regulation.
CC FT DOMAIN 106 115 POLY-GLY.
CC FT DNA_BIND 131 225 FORK-HEAD.
CC FT DOMAIN 252 257 POLY-ALA.
CC FT DOMAIN 265 270 POLY-ALA.
CC FT DOMAIN 275 281 POLY-ALA.
CC FT DOMAIN 380 399 POLY-GLY.
CC FT DOMAIN 447 457 POLY-ALA.
CC SQ SEQUENCE 465 AA; 47092 MW; 6FB5B3D8C7564D CRC64;

Query Match 21.5%; Score 55.5; DB 1; Length 465;
Best Local Similarity 38.6%; Pred. No. 21;
Matches 17; Conservative 3; Mismatches 11; Indels 13; Gaps 2;

QY 7 RCGHVVGLAGLEQLRLA-----EHHPGQREPMMSGCK 43
DB 343 GRCGGHNV-----AHQVRAQCAAVVQHRHHRGIRAPGSAG 380

RESULT 15
ALC_STRCO STANDARD; PRT; 376 AA.
AC Q9RKU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative allantoicase (EC 3.5.3.4) (Allantoate amidinohydrolase).
GN SC06248 OR SCMH10.13 OR STRH10.13.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces coelicolor.
CC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RX Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challie G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabdinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).",
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: Allantoate + H(2)O = (-)-ureidoglycolate +
CC urea.
CC -1- PATHWAY: Degradation of allantoate (purine catabolism); second
CC step.
CC -1- SIMILARITY: Belongs to the allantoicase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AL39126; CAB60167.1; -
DR InterPro; IPR005164; Allantoicase.
DR Pfam; PF03561; Allantoicase; 2.
KW Hypothetical protein; Hydrolase; Purine metabolism; Complete proteome.
SQ SEQUENCE 376 AA; 41207 MW; 2247C0B2300CA29 CRC64;

Query Match 20.9%; Score 54; DB 1; Length 376;
Best Local Similarity 44.4%; Pred. No. 26;
Matches 12; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

QY 10 GGHVEGLAGLEEQ-LRRLRHHPQG 34
DB 171 GGHANGFAVSAEDQRFTHLRKQHPDG 197

Search completed: July 15, 2004, 20:34:28
Job time : 9.84706 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:55 ; Search time 36.4941 Seconds
(without alignments)
406.349 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258

Sequence: 1 MDEKTTGMRGHHVEGLAGE.....LEHHPGQGRPMMSGCKLG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.5	28.5	554	16 Q9A9H1	Q9A9H1 caulobacter
2	66.5	25.8	739	16 Q82BY9	Q82BY9 streptomyces
3	65.5	25.4	189	2 Q8KRD7	Q8KRD7 rhodococcus
4	65.5	25.4	528	16 Q9A365	Q9A365 caulobacter
5	62	24.0	792	16 Q9RDD1	Q9RDD1 streptomyces
6	61	23.6	429	16 Q9K8S7	Q9K8S7 bacillus ha
7	61	23.6	773	2 Q33541	Q33541 endosymbion
8	61	23.6	1161	16 Q9L0L0	Q9L0L0 streptomyces
9	61	23.6	1181	16 Q82DQ5	Q82DQ5 streptomyces
10	60	23.3	377	16 Q9RKG2	Q9RKG2 streptomyces
11	60	23.3	420	16 Q8D3W1	Q8D3W1 vibrio vuln
12	60	23.3	588	16 Q8DH85	Q8DH85 synecchococcus
13	59.5	23.1	342	16 Q8KAY8	Q8KAY8 escherichia
14	59	22.9	139	4 Q96MD1	Q96MD1 homo sapien
15	59	22.9	149	10 Q8L5B8	Q8L5B8 oryza sativ
16	59	22.9	229	16 Q7U554	Q7U554 synecchococcus

17	59	22.9	281	10 Q94JG6	Q94JG6 oryza sativ
18	59	22.9	320	16 Q82ZC6	Q82ZC6 enterococcus
19	59	22.9	342	16 Q83RD8	Q83RD8 shigella fl
20	59	22.9	976	5 Q9U158	Q9U158 leishmania
21	59	22.9	1083	3 Q94189	Q94189 cryptococcus
22	59	22.9	1974	5 Q21000	Q21000 caenorhabditis
23	59	22.9	2360	3 Q94188	Q94188 cryptococcus
24	58.5	22.7	274	16 Q8P4G4	Q8P4G4 xanthomonas
25	58.5	22.7	302	16 Q9KPY2	Q9KPY2 vibrio chol
26	58.5	22.7	1206	10 Q7XNE8	Q7XNE8 oryza sativ
27	58	22.5	333	4 Q8IU04	Q8IU04 homo sapien
28	58	22.5	491	2 Q9ZGD9	Q9ZGD9 streptomyces
29	57.5	22.3	303	16 Q8D9M7	Q8D9M7 vibrio vuln
30	57.5	22.3	440	16 Q8K692	Q8K692 streptococcus
31	57.5	22.3	442	16 Q879H3	Q879H3 streptococcus
32	57.5	22.3	726	5 Q8KX55	Q8KX55 caenorhabditis
33	57.5	22.3	743	5 Q9N313	Q9N313 caenorhabditis
34	57.5	22.3	4001	5 Q8WRQ7	Q8WRQ7 drosophila
35	57.5	22.3	4001	5 Q9VC48	Q9VC48 drosophila
36	57	22.1	93	10 Q8LMT6	Q8LMT6 oryza sativ
37	57	22.1	338	2 Q53739	Q53739 streptomyces
38	57	22.1	381	2 Q83YE1	Q83YE1 streptomyces
39	57	22.1	405	2 Q7X372	Q7X372 uncultured
40	57	22.1	473	2 Q9Z405	Q9Z405 pseudomonas
41	57	22.1	487	13 Q8JGR1	Q8JGR1 brachydanio
42	57	22.1	487	13 Q7ZU28	Q7ZU28 brachydanio
43	57	22.1	489	16 Q8FLU7	Q8FLU7 corynebacter
44	57	22.1	495	2 Q50469	Q50469 mycobacteri
45	57	22.1	496	16 P96284	P96284 mycobacteri

ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	554 AA.
Q9A9H1			
AC Q9A9H1	01-JUN-2001 (TRMBLrel. 17, Created)		
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)			
DT 01-JUN-2003 (TRMBLrel. 24, Last annotation update)			
DE Hypothetical protein CCI017.			
GN CCI017.			
OS Caulobacter crescentus.			
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;			
OC Caulobacteriaceae; Caulobacter.			
OC NCBI_Taxid=155892;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 19089 / CB15;			
RX MEDLINE=21173698; PubMed=11259647;			
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA Bisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,			
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,			
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,			
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,			
RA Usterback T., Tran K., Wolf A., Vamathevan J., Ermlaeva M., White O.,			
RA Salzberg S.L., Venter D.C., Shapiro L., Fraser C.M.;			
RT "Complete genome sequence of Caulobacter crescentus.";			
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			
DR EMBL; AB005780; AAK21001.1; -			
DR PIR; E87375; E87375.			
DR TIGR; CCI017; -			
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.			
DR InterPro; IPR000051; SAM bind.			
KW Hypothetical protein; Complete proteome.			
SQ SEQUENCE 554 AA; 61466 MW; 030FF1804A7CDE80 CRC64;			
Query Match	28.5%	Score 73.5;	DB 16; Length 554;
Best Local Similarity	35.4%	Pred. No. 1.4;	
Matches 17; Conservative	6;	Mismatches 20;	Indels 5; Gaps 1;
QY 5 TTGMRGHHVE-----GLAGELGQLRARLEHHHPGQGRPMMSGCKLG 47			

Db 45 STEMSAGYTDVNYTGYGELNPLRCRLPLTLVGHNAKINACELG 92

RESULT 2

OQ2BY9

ID OQ2BY9 PRELIMINARY; PRT; 739 AA.

AC OQ2BY9; 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Putative regulatory protein.

GN SAV5565.

OS Streptomyces avermitilis.

OC Bacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=33903;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=21477403; PubMed=11572948;

RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,

RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.

RT "Genome sequence of an industrial microorganism Streptomyces

RT avermitilis: deducing the ability of producing secondary

RT metabolites.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN [12]

RN SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=22608306; PubMed=12692562;

RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

RA Sakaki Y., Hattori M., Omura S.;

RT "Complete genome sequence and comparative analysis of the industrial

RT microorganism Streptomyces avermitilis.";

RT Nat. Biotechnol. 21:526-531(2003).

DR EMBL; AP005043; BAC73277.1; -.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003824; F:catalytic activity; IEA.

DR InterPro; IPR003594; ATPbind_ATPase.

DR InterPro; IPR001932; PP2C-like.

DR Pfam; PF02518; HATPase_c; 1.

DR SMART; SM00331; PP2C_SIG; 1.

DR Complete proteome.

SQ SEQUENCE 739 AA; 77478 MW; 63853F93EC67A2B6 CRC64;

SQ Query Match 25.8%; Score 66.5; DB 16; Length 739;

SQ Best Local Similarity 32.6%; Pred. No. 14;

SQ Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;

QY 1 MBEKTTGARG-----GHVVGGLAGLEQLRLRLEHHP 32

Db 341 LDEATGMRGPETGMAPGPRLARVWVHSGENRIEDLRRALRRHP 383

RESULT 3

OQ8KD7

ID OQ8KD7 PRELIMINARY; PRT; 189 AA.

AC OQ8KD7; 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Transcription repressor KfrA-like protein.

OS Rhodococcus erythropolis.

OC Bacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Nocardiaceae; Rhodococcus.

OX NCBI_TaxID=1833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MP50;

RX MEDLINE=22083458; PubMed=12089004;

RA Troelt S., Burger S., Calaminius C., Stolz A.;

RT "Cloning and Heterologous Expression of an Enantioselective Amidase
RT from Rhodococcus erythropolis Strain MP50."
RT Appl. Environ. Microbiol. 68:3279-3286(2002).
DR EMBL; AY026386; AAK11725.1; -.
SQ SEQUENCE 189 AA; 20101 MW; 4B8B5D0A093FFFD CRC64;

SQ Query Match 25.4%; Score 65.5; DB 2; Length 189;

SQ Best Local Similarity 45.5%; Pred. No. 4;

SQ Matches 20; Conservative 2; Mismatches 13; Indels 9; Gaps 1;

QY 3 EKTGMRGHHV-----VEGLAGLEQLRLRLEHHPGORE 37

Db 104 EQATERRGAVTARDRALGVEGLRVELEQRGLQETARQDARE 147

RESULT 4

OQ9A65

ID OQ9A65 PRELIMINARY; PRT; 528 AA.

AC OQ9A65; 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Hypothetical protein CC3341.

GN CC3341.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;

OC Caulobacteraceae; Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CBA15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke D.H.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gaim M.L., Hatt D.H.,

RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathavan J., Enmolava M., White O.,

RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005995; AAK25303.1; -.

DR PIR; C87663; C87663.

DR TIGR; CC3341; -.

DR GO; GO:0008168; F:methyltransferase activity; IEA.

DR InterPro; IPR01601; Methyltransf.

DR Hypothetical protein; Complete proteome.

SQ SEQUENCE 528 AA; 57050 MW; E18A1F4DA1785D5 CRC64;

SQ Query Match 25.4%; Score 65.5; DB 16; Length 528;

SQ Best Local Similarity 37.8%; Pred. No. 12;

SQ Matches 17; Conservative 5; Mismatches 18; Indels 5; Gaps 1;

QY 8 WRGCHVE-----GLAGLEQLRLRLEHHPGOREPMMGGCKLQ 47

Db 23 MSAGVIVETGYTGCGYRPLNPVRALPLLAGFPAKTIETACELG 67

RESULT 5

OQ9DD1

ID OQ9DD1 PRELIMINARY; PRT; 792 AA.

AC OQ9DD1; 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Hypothetical protein SC02560.

GN SC02560 OR SC02560.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

```
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Rendenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warin T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939113; CAB6238.1; -
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR003018; GAP.
DR InterPro: IPR001932; PP2C-like.
DR Pfam: PF02518; HATPase_c; 1.
DR SMART: SM00065; GAP; 1.
DR SMART: SM00387; HATPase_c; 1.
DR SMART: SM00331; PP2C_SIG; 1.
DR SMART: SM00331; PP2C_SIG; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 792 AA; 81804 MW; EFBVCCE66852D3D CRC64;
SQ
Query Match 24.0%; Score 62; DB 16; Length 792;
Best Local Similarity 40.0%; Pred. No. 53;
Matches 14; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
QY 8 WRCGHVEGLAGLEQLRRLRLEHHPOGQREPMMSG 42
DB 422 WHSGE-----GHLDLRALRLEKDPFRDPMRSG 450
RESULT 6
O9K8S7 PRELIMINARY; PRT; 429 AA.
ID O9K8S7
AC O9K8S7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Maltose/malodextrin transport system (Maltose/malodextrin-binding
protein).
GN BH2926.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maesu N.,
```

```
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001517; BAB06645.1; -
DR PIR: F84015; F84015.
DR HSSP: P02926; 1IUD.
DR GO: GO:0005363; P:maltose transporter activity; IEA.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0015768; P:maltose transport; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR006060; Maltose BP.
DR InterPro: IPR000437; Prok_lipoprct_S.
DR InterPro: IPR006059; SBP_dac_1.
DR InterPro: IPR006061; SBP_dcm1.
DR Pfam: PF01547; SBP_dac_1; 1.
DR PRINTS: PR00181; MALTOSBP.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01037; SBP_BACTERIAL_1; 1.
KW Complete proteome.
SQ SEQUENCE 429 AA; 46536 MW; 5B17521A0FA19297 CRC64;
QY 10 GCHVVEGLAGLEQLRRLRLEHHPOGQREPMMSGCKLG 47
DB 122 GDVYQLGLAELRLELPQLEGYFEGADALNVEGAQLG 159
RESULT 7
O33541 PRELIMINARY; PRT; 773 AA.
ID O33541
AC O33541;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Sensor protein RsaA.
GN RsaA.
OS endosymbiont of Rickettsia pachyptila.
OC Bacteria; Proteobacteria; Gammaproteobacteria;
OC Sulfur-oxidizing symbionts.
OX NCBI_TaxID=54396;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97438512; PubMed=9293000;
RA Hughes D.S., Felbeck H., Stein J.L.;
RT "A histidine protein kinase homolog from the endosymbiont of the
hyperthermophilic vent tubeworm Rickettsia pachyptila.";
RL Appl. Environ. Microbiol. 63:3494-3498(1997).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
KINASES.
DR EMBL: U93704; AAB7131.1; -
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0016301; F:kinase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0000156; F:two-component response regulator activity; IEA.
DR GO: GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO: GO:0007600; P:two-component signal transduction system (p. . .); IEA.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR003461; His_kina_N.
DR InterPro: IPR001789; Response_reg.
DR Pfam: PF02518; HATPase_c; 1.
DR Pfam: PF00512; HSKA_1.
DR Pfam: PF00072; response_reg; 1.
DR PRINTS: PR00344; BCTR1SENSOR.
```

DR ProDom; PD000039; Response reg; 1.
 DR SMART; SM00387; HATPase C; 1.
 DR SMART; SM00388; HATPase C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS0109; HIS_KIN; 1.
 DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
 DR Kinase; Phosphorylation; Sensory transduction; Transferase.
 SQ SEQUENCE 773 AA; 87086 MW; 2C5643B64F373827 CRC64;
 Query Match 23.6%; Score 61; DB 2; Length 773;
 Best Local Similarity 43.3%; Pred. No. 68;
 Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 8 MRGSHVEGLAGELGRLARLEHHPQGGRE 37
 DB 346 MRGSHVEGLAGELGRLARLEHHPQGGRE 375
 RESULT 8
 ID 091010 PRELIMINARY; PRT; 1161 AA.
 AC 091010;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE DNA-directed RNA polymerase beta chain.
 GN RPOB OR SCO4654 OR SCD82.26.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 NCBI_TaxID=1902;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kiese R.H.M., Denepate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese R.H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kiese R.H., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierhorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)." ;
 RL Nature 417:141-147(2002).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1

CC BETA' CHAIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 CC EMBL; AL393121; CAB77428.1; -.
 DR HSP; Q9K07; IROM.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:006350; P:transcription; IEA.
 DR InterPro; IPR007121; RNA_pol_B.
 DR InterPro; IPR007644; RNA_pol_Rpb2_1.
 DR InterPro; IPR007642; RNA_pol_Rpb2_2.
 DR InterPro; IPR007645; RNA_pol_Rpb2_3.
 DR InterPro; IPR007120; RNA_pol_Rpb2_6.
 DR InterPro; IPR007641; RNA_pol_Rpb2_7.
 DR Pfam; PF04563; RNA_pol_Rpb2_1; 2.
 DR Pfam; PF04561; RNA_pol_Rpb2_2; 2.
 DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE; PS01166; RNA_POL_BETA; 1.
 KW DNA-directed RNA polymerase; Transcription; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 1161 AA; 128494 MW; A9C27B928BE191A8 CRC64;
 Query Match 23.6%; Score 61; DB 16; Length 1161;
 Best Local Similarity 39.5%; Pred. No. 11e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 2;
 QY 4 KTTGMRGSHVEGLAGELGRLARLEHHPQGGREPM 40
 DB 227 KALGWTTEQLIEEF-GEYESMRATLEKDHOTGGODALL 263
 RESULT 9
 ID 082D05 PRELIMINARY; PRT; 1181 AA.
 AC 082D05;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative RNA polymerase beta subunit.
 GN RPOB OR SAV4914.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hatton M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hatton M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis." ;
 RL Nat. Biotechnol. 21:526-531(2003).
 CC EMBL; AP005040; BAC72626.1; -.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR InterPro; IPR007121; RNA_pol_B.
 DR InterPro; IPR007644; RNA_pol_Rpb2_1.
 DR InterPro; IPR007642; RNA_pol_Rpb2_2.

DR InterPro: IPR007645; RNA_pol_Rpb2_3.
 DR InterPro: IPR007120; RNA_pol_Rpb2_6.
 DR InterPro: IPR007644; RNA_pol_Rpb2_7.
 DR Pfam: PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam: PF04561; RNA_pol_Rpb2_2; 1.
 DR Pfam: PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam: PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam: PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE: PS01166; RNA_POL_BETA; 1.
 DR Complete proteome.
 SQ SEQUENCE 1181 AA; 130439 MW; ABBF77ACC879ED8B CRC64;

Query Match 23.6%; Score 61; DB 16; Length 1181;
 Best Local Similarity 39.5%; Pred. No. 1.1e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 2;

QY 4 KTTGWRGHHVEGLAGELEQLRLRL-HPDQGRPEPM 40
 DB 247 KALGWTTEQLLEEF-GEYESMRATLEKDTQGDML 283

RESULT 10
 ID Q9RKG2 PRELIMINARY; PRT; 377 AA.
 AC Q9RKG2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative isomerase.
 GN SC03475 OR SCE55.11C OR SCE55.11C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Streptomyces; Streptomyces.
 NC NCB1_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cardeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RL "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warten T., Wierhorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL: AL339116; CAB6179.1; -.
 DR HSSP: P42206; 1BQG.
 DR GO: GO:0016523; F:isomerase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.

DR InterPro: IPR001354; MR_MLE.
 DR Pfam: PF01188; MR_MLE_1.
 DR Pfam: PF02746; MR_MLE_N; 1.
 DR PROSITE: PS00908; MR_MLE_1; 1.
 DR PROSITE: PS00909; MR_MLE_2; 1.
 DR Isomerase; Complete proteome.
 SQ SEQUENCE 377 AA; 40887 MW; 8787D44A30985E91 CRC64;

Query Match 23.3%; Score 60; DB 16; Length 377;
 Best Local Similarity 37.9%; Pred. No. 41;
 Matches 22; Conservative 4; Mismatches 18; Indels 14; Gaps 4;

QY 2 DEKTTGWRGHHVEGLA-----GELRLRLA-----RLHHPOGQ-REPMMSGGCKL 46
 DB 23 DKGVIGW-GEAVVEQARVERMIDLAELAYLRGRPRRIHHWQGMKGFYRGAVL 79

RESULT 11
 ID Q8D3W1 PRELIMINARY; PRT; 420 AA.
 AC Q8D3W1;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 125, Last annotation update)
 DE ABC-type protease/lipase transport system, ATPase and permease component.
 GN VV21572.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 NC NCB1_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RL "Complete genome sequence of Vibrio vulnificus CMCP6";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016813; AA008435.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
 DR GO: GO:000166; F:nucleotide binding; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:006810; P:transport; IEA.
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR003439; ABC transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 DR Protease; Complete proteome.
 SQ SEQUENCE 420 AA; 45496 MW; 98A77D0FA3E3A51B CRC64;

Query Match 23.3%; Score 60; DB 16; Length 420;
 Best Local Similarity 21.1%; Pred. No. 46;
 Matches 16; Conservative 13; Mismatches 17; Indels 30; Gaps 3;

QY 1 MDEKTTGWRGHHVEGLAGELEQLRLRL-HPDQGRPEPM 38
 DB 137 MGAATTTMSAALASAYGRUKESDKQVDBLNMPLRPRRGHPELANVSYHNP-GVAEP 195

QY 39 MMSG-----GCKLG 47
 DB 196 VLSGAQFSPGCAVG 211

RESULT 12
 ID Q8DH85 PRELIMINARY; PRT; 588 AA.
 AC Q8DH85;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

DT 01-OCT-2003 (TReMBLrel. 25, last annotation update)
DE Penicillin-binding protein.
GN TLR2074.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_Taxid=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1.
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005376; BAC09626.1; -.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR Complete proteome.
SQ SEQUENCE 588 AA; 65048 MW; 9E24A8544A102719 CRC64;

Query Match 23.3%; Score 60; DB 16; Length 588;
Best Local Similarity 57.1%; Pred. No. 67;
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 12 HVVEGLAGELEQRLARLEHHP 32
Db 442 HVVDGLYDENGQLQKRLHHP 462

RESULT 13
Q8XAY8 PRELIMINARY; PRT; 342 AA.
AC Q8XAY8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
DE Putative transport system permease protein.
GN YDEY OR Z2191 OR ECS2121.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=1120651;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Tobe T.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AB005353; AAG56252.1; -.

```

```

DR EMBL; AP002557; BAB35544.1; -.
DR PIR; A90894; A90894.
DR PIR; H85723; H85723.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
DR Complete proteome.
SQ SEQUENCE 342 AA; 36365 MW; 6953A23C07C305A8 CRC64;

Query Match 23.1%; Score 59.5; DB 16; Length 342;
Best Local Similarity 35.0%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 10; Indels 11; Gaps 1;

Qy 8 WRGHHVGLAGELEQRLARLEHHPGQREPMMSGCKLG 47
Db 135 WTGGRWIEGLPAELKQLSA-----PULLGVSAIG 163

RESULT 14
Q96MD1 PRELIMINARY; PRT; 139 AA.
AC Q96MD1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)
DE Hypothetical protein FLJ32575.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaetsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK051137; BAB71369.1; -.
DR Hypothetical protein.
SQ SEQUENCE 139 AA; 15014 MW; 3A59D86197B48F19 CRC64;

Query Match 22.9%; Score 59; DB 4; Length 139;
Best Local Similarity 44.8%; Pred. No. 18;
Matches 13; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Qy 3 EKTGWGGRGHVVEGLAGELEQRLARLEHHP 31
Db 92 ERATGWRLSHVEGSGGLQATSCR--RH 118

RESULT 15
Q8LSB8 PRELIMINARY; PRT; 149 AA.
AC Q8LSB8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, last annotation update)
DE P0671D01.7 protein (P0445H04.30 protein).
GN P0671D01.7 OR P0445H04.30.
OS Oryza sativa [Japonica cultivar-group].
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrharioideae; Oryzaeae; Oryza.
OX NCBI_Taxid=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;

```


RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0671D01."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0445H04."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003284; BAB91731.1; -
DR EMBL; AP004364; BAC06318.1; -
DR Gramene; O815B8; -
SQ SEQUENCE 149 AA; 16716 MW; 0081630FBE03018D CRC64;

Query Match 22.9%; Score 59; DB 10; Length 149;
Best Local Similarity 43.3%; Pred. No. 19;
Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 17 LAGLEQLRLRLHHHPGQREPMMSGCKL 46
DB 77 LAGDLSTYRESIRFHPPAPPAKGGGL 106

Search completed: July 15, 2004, 20:35:47
Job time : 38.4941 secs

is Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:29:44 ; Search time 53.0824 Seconds
(without alignments)
250.172 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDKKTGMKGHVGEGLAGE.....LEHHPGQGRPMWGGCKLKG 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29aug04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	47	5	ABG67079 Streptavi
2	258	100.0	47	6	ABUS7601 Synthetic
3	243	94.2	47	5	ABG67080 Streptavi
4	243	94.2	47	6	ABUS7602 Streptavi
5	241	93.4	61	5	ABG67078 Streptavi
6	241	93.4	61	6	ABUS7600 Synthetic
7	234	90.7	75	5	ABG67077 Streptavi
8	234	90.7	75	6	ABUS7599 Synthetic
9	227	88.0	89	5	ABG67076 Streptavi
10	227	88.0	89	6	ABUS7598 Synthetic
11	214	82.9	101	5	ABG67075 Streptavi
12	214	82.9	101	6	ABUS7597 Streptavi
13	211	81.8	101	5	ABG67073 Streptavi
14	211	81.8	101	6	ABUS7595 Streptavi
15	208	80.6	38	5	ABG67087 Streptavi
16	208	80.6	38	6	ABUS7609 Streptavi
17	208	80.6	426	5	ABG67091 Streptavi
18	208	80.6	426	6	ABUS7613 Streptavi
19	208	80.6	479	5	ABG67089 Streptavi
20	208	80.6	479	6	ABUS7611 Streptavi
21	199	77.1	39	5	ABG67081 Streptavi
22	199	77.1	39	6	ABUS7603 Streptavi
23	152	58.9	29	5	ABG67083 Streptavi
24	152	58.9	29	6	ABUS7605 Synthetic
25	133	51.6	89	5	ABG67082 Streptavi

26	133	51.6	89	6	ABUS7604 Synthetic
27	94	36.4	44	5	ABG67086 Streptavi
28	88	34.1	44	6	ABUS7608 Streptavi
29	70	27.1	102	5	ABG67065 Streptavi
30	70	27.1	102	6	ABUS7587 Streptavi
31	68	26.4	101	5	ABG67066 Streptavi
32	68	26.4	101	6	ABUS7588 Streptavi
33	63	24.4	145	5	ABP09375 Human ORF
34	62	24.0	164	4	ABG63519 Human gas
35	62	24.0	336	6	ABH67778 Photorhab
36	61	23.6	288	6	ABR96152 Human NOV
37	60.5	23.4	741	4	ABG25060 Novel hum
38	59.5	23.1	84	5	ABP34332 Human hel
39	59.5	23.1	342	3	ABR15903 E. coli P
40	59.5	23.1	342	6	ABU14680 Protein e
41	59.5	23.1	538	4	ABG25106 Novel hum
42	59	22.9	139	6	ADA55387 Human pro
43	59	22.9	140	2	AAW40203 FRAP homo
44	59	22.9	147	4	AAH19547 Peptide #
45	59	22.9	147	4	ABR39099 Peptide #

ALIGNMENTS

RESULT 1	ABG67079 standard, peptide, 47 AA.
ID	ABG67079
XX	AC ABG67079;
XX	24-SEP-2002 (first entry)
XX	Streptavidin-binding peptide SB19 truncation mutant #5.
XX	Streptavidin-binding peptide; mutant; mutein.
KW	Escherichia coli.
OS	WO200238580-A1.
XX	16-MAY-2002.
XX	31-OCT-2000; 2000WO-US041717.
XX	31-OCT-2000; 2000WO-US041717.
XX	(GENO) GEN HOSPITAL CORP.
XX	Szostak JW, Wilson DS, Keefe AD;
XX	WPI; 2002-500155/53.
XX	Novel peptide with high affinity for streptavidin, is expressed as part of fusion protein to facilitate detection, quantitation and purification of desired protein.
XX	Claim 9; Fig 5; 56pp; English.
XX	The invention describes a peptide (I) which binds streptavidin with a dissociation constant less than 10 mM or 23 nM, where the amino acid sequence of (I) does not contain an HPQ, HPM, HPN or HOP motif, is not disulfide bonded or cyclised. A fusion protein comprising a protein of interest covalently linked to (I) is useful for purifying a desired protein from a sample by contacting the sample with streptavidin under conditions that allow complex formation between the fusion protein and the streptavidin, isolating the complex and recovering the fusion protein, and thus purifying the desired protein from the sample. The invention describes a method for producing a streptavidin-binding fusion protein. This sequence represents a truncation mutant of the streptavidin-binding peptide SB-19 isolated from Escherichia coli using a method described in the specification

SQ Sequence 47 AA;

Query Match 100.0%; Score 258; DB 5; Length 47;
 Best Local Similarity 100.0%; Pred. No. 2,3e-28;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47
 DB 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47

RESULT 2
 ABUS7601
 ID ABUS7601 standard; peptide; 47 AA.

AC ABUS7601;

DT 09-APR-2003 (first entry)

DE Synthetic peptide SB19 C-terminal deletion mutant, C4.

XX Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.

XX Synthetic.

OS US200215578-A1.

PN 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

PA (SZOS/) SZOSTAK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szostak JW, Wilson DS, Keefe AD;

XX WPI; 2003-182639/18.

PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.

PS Claim 14; Fig 5; 36pp; English.

XX The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulphide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX Sequence 47 AA;

Query Match 100.0%; Score 258; DB 6; Length 47;

Best Local Similarity 100.0%; Pred. No. 2,3e-28;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47
 DB 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47

RESULT 3
 ABG67080
 ID ABG67080 standard; peptide; 47 AA.

AC ABG67080;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19 truncation mutant #6.

XX Streptavidin-binding peptide; mutant; mutein.
 KW Escherichia coli.

OS Escherichia coli.

FN WO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US041717.

PR 31-OCT-2000; 2000WO-US041717.

PA (GEHO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

XX WPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.

PS Claim 9; Fig 5; 56pp; English.

XX The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HPO, HPN, HPN or HQP motif, is not
 CC disulphide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a protein of
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX Sequence 47 AA;

Query Match 94.2%; Score 243; DB 5; Length 47;
 Best Local Similarity 95.7%; Pred. No. 2,8e-26;
 Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47
 DB 1 MDEKTTGMRGHHVVEGLAGLEQLRARLEHHHPQGRPEPMMSGCKLG 47

RESULT 4
 ABUS7602
 ID ABUS7602 standard; peptide; 47 AA.

AC ABUS7602;

XX

DT 09-APR-2003 (first entry)
 XX Streptavidin-binding synthetic peptide SB19 point mutant M1.
 DE
 XX
 XX Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 XX Synthetic.
 OS
 PN US2002155578-A1.
 XX
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 XX (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.
 XX
 CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQ, HPW, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX
 SQ Sequence 47 AA;
 XX
 Query Match 94.2%; Score 243; DB 6; Length 47;
 Best Local Similarity 95.7%; Pred. No. 2.8e-26;
 Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGELRLRLRHHHPQGRPPMSSGGCKLG 47
 DB 1 MDEKTTGMRGHHVVEGLAGELRLRLRHHHPQGRPPMSSGGCKLG 47

RESULT 5
 ABG67078
 ID ABG67078 standard; peptide; 61 AA.
 XX
 AC ABG67078;
 XX
 DT 24-SEP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #4.

XX
 KW Streptavidin-binding peptide; mutant; mutein.
 XX
 XX Escherichia coli.
 OS
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX
 PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2002-500155/53.
 XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 PS Claim 9; Fig 5; 56pp; English.
 XX
 CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 nM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQ, HPW, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX
 SQ Sequence 61 AA;
 XX
 Query Match 93.4%; Score 241; DB 5; Length 61;
 Best Local Similarity 77.0%; Pred. No. 7.3e-26;
 Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 MDEKTTGMRGHHVVEGLAGELRLRLRHHHPQGRPPMSSGGCKL 46
 DB 1 MDEKTTGMRGHHVVEGLAGELRLRLRHHHPQGRPPMSSGGCKL 60
 QY 47 G 47
 DB 61 G 61

RESULT 6
 ABUS7600
 ID ABUS7600 standard; peptide; 61 AA.
 XX
 AC ABUS7600;
 XX
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C3.
 XX
 KW Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX

PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.
 XX
 CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPM, HPN or HOP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX
 SQ Sequence 61 AA;
 Query Match 93.4%; Score 241; DB 6; Length 61;
 Best Local Similarity 77.0%; Pred. No. 7.3e-26;
 Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGELEQLRLRLHHHPGQREP-----WMSGGCKL 46
 DB 1 MDEKTTGMRGSHVVEGLAGELEQLRLRLHHHPGQREPVEVEDVDEGLVQMSGGCKL 60
 QY 47 G 47
 DB 61 G 61
 RESULT 7
 ABG67077
 ID ABG67077 standard; peptide; 75 AA.
 XX
 AC ABG67077;
 XX
 DT 24-SBP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #3.
 XX
 KW Streptavidin-binding peptide; mutant; mutein.
 XX
 OS Escherichia coli.
 XX
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX

PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2002-500155/53.
 XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 PS Claim 9; Fig 5; 56pp; English.
 XX
 CC The invention describes a peptide (1) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (1) does not contain an HPO, HPM, HPN or HOP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (1) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX
 SQ Sequence 75 AA;
 Query Match 90.7%; Score 234; DB 5; Length 75;
 Best Local Similarity 62.7%; Pred. No. 8.7e-25;
 Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGELEQLRLRLHHHPGQREP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAGELEQLRLRLHHHPGQREPVEVEDVDEGLVQDLHGAVAG 60
 QY 39 -----WMSGGCKLG 47
 DB 61 ILDPVEMMSGGCKLG 75
 RESULT 8
 ABUS7599
 ID ABUS7599 standard; peptide; 75 AA.
 XX
 AC ABUS7599;
 XX
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C2.
 XX
 KW Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;

XX DR WPI; 2003-182639/18.
 XX PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 XX interest.
 XX PS Claim 14; Fig 5; 36pp; English.
 XX CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQ, HPM, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX SQ Sequence 75 AA;
 Query Match 90.7%; Score 234; DB 6; Length 75;
 Best Local Similarity 62.7%; Pred. No. 8.7e-25;
 Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHPOGOREP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHPOGOREPVEEDVDEGLVODLHGAVG 60
 QY 39 -----MMSGGCKLG 47
 DB 61 LLDPEVEMSGGCKLG 75
 RESULT 9
 ABG67076
 ID ABG67076 standard; peptide; 89 AA.
 AC
 XX ABG67076;
 DT 24-SEP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #2.
 XX
 KM Streptavidin-binding peptide; mutant; mutein.
 XX
 OS Escherichia coli.
 XX
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX
 PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2002-500155/53.

XX PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 XX PS Claim 9; Fig 5; 56pp; English.
 XX CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQ, HPM, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX SQ Sequence 89 AA;
 Query Match 88.0%; Score 227; DB 5; Length 89;
 Best Local Similarity 52.8%; Pred. No. 1e-23;
 Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHPOGOREP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHPOGOREPVEEDVDEGLVODLHGAVG 60
 QY 39 -----MMSGGCKLG 47
 DB 61 LLDPEVKLITDWFKKFKVNSMMSGGCKLG 89
 RESULT 10
 ABUS7598
 ID ABUS7598 standard; peptide; 89 AA.
 AC
 XX ABUS7598;
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C1.
 XX
 KM Streptavidin-binding peptide; mRNA display; peptide library;
 XX fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 XX (WILS/) WILSON D S.
 XX (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2003-182639/18.
 XX
 XX PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 XX interest.
 XX PS Claim 14; Fig 5; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulfide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPW, HPN or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations

SQ Sequence 89 AA;

Query Match 88.0%; Score 227; DB 6; Length 89;
Best Local Similarity 52.8%; Pred. No. 1e-23;
Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREP----- 38
1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREPLVQVEVDVDEGLVODLHGAVAG 60

DB 39 -----MMSGCKLG 47
61 LLDPEVKLLTDMFKFKRNVSMDCKMTFYLEMYDMSGCKLG 89

RESULT 11
ABG67075
ID ABG67075 standard; peptide; 101 AA.

AC ABG67075;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19 truncation mutant #1.

KW Streptavidin-binding peptide; mutant; mutein.

OS Escherichia coli.

OS Synthetic.

XX MO200238580-A1.

XX 16-MAY-2002.

XX 31-OCT-2000; 2000WO-US041717.

XX 31-OCT-2000; 2000WO-US041717.

XX (GENO) GEN HOSPITAL CORP.

XX Szostak JW, Wilson DS, Keefe AD;

XX WPI; 2002-500155/53.

XX Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

PS Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a

CC dissociation constant less than 10 μ M or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPO, HPW, HPN or HOP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a truncation mutant of the streptavidin
CC -binding peptide SB-19 isolated from Escherichia coli using a method
CC described in the specification

SQ Sequence 101 AA;

Query Match 82.9%; Score 214; DB 5; Length 101;
Best Local Similarity 44.6%; Pred. No. 7.7e-22;
Matches 45; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREP----- 40
1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREPLVQVEVDVDEGLVODLHGAVAG 60

DB 41 -----SQGCKLG 47
61 LLDPEVKLLTDMFKFKRNVSMDCKMTFYLEMYDMSGCKLG 101

RESULT 12
ABU57597
ID ABU57597 standard; peptide; 101 AA.

AC ABU57597;

DT 09-APR-2003 (first entry)

DE Streptavidin-binding synthetic peptide SB19 full length peptide.

KW Streptavidin-binding peptide; mRNA display; peptide library;

XX fusion protein.

XX Synthetic.

XX US2002155578-A1.

XX 24-OCT-2002.

XX 31-OCT-2001; 2001US-00004381.

XX 31-OCT-2000; 2000US-0244541P.

XX (SZOS/) SZOSTAK J W.

XX (WILS/) WILSON D S.

XX (KEEF/) KEEFE A D.

XX Szostak JW, Wilson DS, Keefe AD;

XX WPI; 2003-182639/18.

XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.

PS Claim 14; Fig 5; 36pp; English.

XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulfide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPW, HPN or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a

CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a streptavidin binding peptide of the invention
XX

SQ Sequence 101 AA;

Query Match 82.9%; Score 214; DB 6; Length 101;

Best Local Similarity 44.6%; Pred. No. 7.7e-22;

Matches 45; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPM-----40

Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPLVQVEYEDVDEGLVQDLHGAVAG 60

QY 41 -----SGGCKLG 47

Db 61 LLDPEKLTLDWFKKFNVSCKDKMTFYLEMYDMSGCKLG 101

RESULT 13

ABG67073 standard; peptide; 101 AA.

AC ABG67073;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19.

KM Streptavidin-binding peptide.

OS Escherichia coli.

PN W0200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US041717.

PR 31-OCT-2000; 2000WO-US041717.

PA (GEHO) GEN HOSPITAL CORP.

PI Szoestak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

XX Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

PS Claim 9; Fig 3; 56pp; English.

XX The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPO, HPN, HPN or HOP motif, is not
CC disulfide bonded or cyclized. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the complex and recovering the fusion
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a streptavidin-binding peptide isolated
CC from Escherichia coli using a method described in the specification

XX SQ Sequence 101 AA;

Query Match 81.8%; Score 211; DB 5; Length 101;

Best Local Similarity 95.0%; Pred. No. 2e-21;

Matches 38; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPM 40

Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPLV 40

RESULT 14

ABUS7595 standard; peptide; 101 AA.

AC ABUS7595;

DT 09-APR-2003 (first entry)

DE Streptavidin-binding synthetic peptide SB19.

KM Streptavidin-binding peptide; mRNA display; peptide library;

XX fusion protein.

OS Synthetic.

PN US2002155578-A1.

PD 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

PA (SZOS/) SZOSTAK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szoestak JW, Wilson DS, Keefe AD;

DR WPI; 2003-182639/18.

XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.

PS Claim 14; Fig 3; 36pp; English.

XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulfide bonded or cyclized, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPN, HPN or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a streptavidin binding peptide of the invention

SQ Sequence 101 AA;

Query Match 81.8%; Score 211; DB 6; Length 101;

Best Local Similarity 95.0%; Pred. No. 2e-21;

Matches 38; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHHPQGOREPM 40
 |||
 Db 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHHPQGOREPLV 40
 |||

RESULT 15

ABG67087
 ID ABG67087 standard; peptide; 38 AA.

AC ABG67087;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide and affinity tag.

KW Streptavidin-binding peptide.

OS Synthetic.

PN WO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US041717.

PR 31-OCT-2000; 2000WO-US041717.

PA (GENO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
 of fusion protein to facilitate detection, quantitation and purification
 of desired protein.

PS Claim 13; Fig 7A; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HPO, HPN, HPN or HOP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a streptavidin-binding peptide used in
 CC the creation of streptavidin-binding DNA-tagged peptide described in the
 CC invention

SQ Sequence 38 AA;

Query Match 80.6%; Score 208; DB 5; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.7e-21;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHHPQGOREP 38
 |||
 Db 1 MDEKTTGMRGSHVVEGLAGLEQLRLRLHHHPQGOREP 38
 |||

Search completed: July 15, 2004, 20:34:02
 Job time : 55.0824 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:35:51 ; Search time 35.7647 Seconds
(without alignments)
332.096 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208
Sequence: 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLHHHOGQREP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	208	100.0	38	US-10-004-381-35	Sequence 35, Appl
2	208	100.0	47	US-10-004-381-25	Sequence 25, Appl
3	208	100.0	61	US-10-004-381-24	Sequence 24, Appl
4	208	100.0	75	US-10-004-381-23	Sequence 23, Appl
5	208	100.0	89	US-10-004-381-22	Sequence 22, Appl
6	208	100.0	101	US-10-004-381-19	Sequence 19, Appl
7	208	100.0	101	US-10-004-381-21	Sequence 21, Appl
8	208	100.0	426	US-10-004-381-41	Sequence 41, Appl
9	208	100.0	479	US-10-004-381-38	Sequence 38, Appl
10	193	92.8	47	US-10-004-381-26	Sequence 26, Appl
11	149	71.6	39	US-10-004-381-27	Sequence 27, Appl
12	127	61.1	89	US-10-004-381-28	Sequence 28, Appl
13	102	49.0	29	US-10-004-381-29	Sequence 29, Appl
14	67	32.2	102	US-10-004-381-11	Sequence 11, Appl
15	66.5	32.0	739	US-10-156-761-13097	Sequence 13097, A

16	66	31.7	101	13	US-10-004-381-12	Sequence 12, Appl
17	62	29.8	527	12	US-10-425-114-70961	Sequence 70961, A
18	62	29.8	529	12	US-10-425-114-37475	Sequence 37475, A
19	62	29.8	649	12	US-10-425-114-70144	Sequence 70144, A
20	59	28.4	139	15	US-10-094-749-2955	Sequence 2955, Ap
21	59	28.4	140	9	US-09-902-180-6	Sequence 6, Appl1
22	59	28.4	249	12	US-10-112-944-859	Sequence 859, App
23	59	28.4	249	12	US-10-112-944-860	Sequence 860, App
24	59	28.4	249	12	US-10-112-944-861	Sequence 861, App
25	59	28.4	249	12	US-10-112-944-862	Sequence 862, App
26	59	28.4	342	9	US-09-912-020-260	Sequence 260, App
27	59	28.4	342	12	US-10-282-122A-42604	Sequence 42604, A
28	59	28.4	869	16	US-10-437-963-156107	Sequence 156107, A
29	58.5	28.1	175	15	US-10-437-963-147749	Sequence 147749, A
30	58	27.9	288	12	US-10-112-944-410	Sequence 410, App
31	58	27.9	298	12	US-10-423-584-2	Sequence 2, Appl1
32	58	27.9	300	12	US-10-112-944-411	Sequence 411, App
33	58	27.9	306	12	US-10-112-944-413	Sequence 413, App
34	58	27.9	333	12	US-10-112-944-412	Sequence 412, App
35	58	27.9	333	12	US-10-423-582-2	Sequence 2, Appl1
36	58	27.9	333	12	US-10-411-120-69	Sequence 69, Appl
37	58	27.9	333	12	US-10-411-120-97	Sequence 97, Appl
38	58	27.9	1181	14	US-10-156-761-12448	Sequence 12448, A
39	58	27.9	2222	15	US-10-369-493-13923	Sequence 3923, Ap
40	57	27.4	114	12	US-10-425-114-61050	Sequence 61050, A
41	57	27.4	474	9	US-09-738-626-6856	Sequence 6856, Ap
42	56.5	27.2	939	16	US-10-437-963-132193	Sequence 132193, A
43	55.5	26.7	1336	16	US-10-437-963-115864	Sequence 115864, A
44	55.5	26.7	1388	12	US-10-332-089-2	Sequence 2, Appl1
45	55.5	26.7	1388	14	US-10-146-473-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1

US-10-004-381-35

Sequence 35, Application US/10004381

Publication No. US20020155578A1

GENERAL INFORMATION:

APPLICANT: SZOSTAK, JACK W.

APPLICANT: WILSON, DAVID S.

APPLICANT: KEEPE, ANTHONY D.

TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

FILE REFERENCE: 00786/388002

CURRENT APPLICATION NUMBER: US/10/004,381

CURRENT FILING DATE: 2001-10-31

PRIOR APPLICATION NUMBER: US 60/244,541

PRIOR FILING DATE: 2000-10-31

NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 35

LENGTH: 38

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: selected peptide

US-10-004-381-35

Query Match

Best Local Similarity 100.0%; Score 208; DB 13; Length 38;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 MDEKTTGMRGHHVVEGLAGLEQLRLARLHHHOGQREP 38

|||||

1 MDEKTTGMRGHHVVEGLAGLEQLRLARLHHHOGQREP 38

Db

1 MDEKTTGMRGHHVVEGLAGLEQLRLARLHHHOGQREP 38

RESULT 2

US-10-004-381-25

Sequence 25, Application US/10004381

Publication No. US20020155578A1

```
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 25
/ LENGTH: 47
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-25
```

```
Query Match          100.0%; Score 208; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 3,4e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
```

```
RESULT 3
US-10-004-381-24
/ Sequence 24, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 24
/ LENGTH: 61
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-24
```

```
Query Match          100.0%; Score 208; DB 13; Length 61;
Best Local Similarity 100.0%; Pred. No. 4,5e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
```

```
RESULT 4
US-10-004-381-23
/ Sequence 23, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
```

```
/ TITLE OF INVENTION: THEREOF
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 23
/ LENGTH: 75
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-23
```

```
Query Match          100.0%; Score 208; DB 13; Length 75;
Best Local Similarity 100.0%; Pred. No. 5,6e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
```

```
RESULT 5
US-10-004-381-22
/ Sequence 22, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 22
/ LENGTH: 89
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-22
```

```
Query Match          100.0%; Score 208; DB 13; Length 89;
Best Local Similarity 100.0%; Pred. No. 6,7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGGREP 38
```

```
RESULT 6
US-10-004-381-19
/ Sequence 19, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
```

```

; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-19

Query Match          100.0%; Score 208; DB 13; Length 101;
Best Local Similarity 100.0%; Pred. No. 7.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38

RESULT 7
US-10-004-381-21
; Sequence 21, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: WILSON, JACK W.
; APPLICANT: SZOSTAK, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-21

Query Match          100.0%; Score 208; DB 13; Length 101;
Best Local Similarity 100.0%; Pred. No. 7.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38

RESULT 8
US-10-004-381-41
; Sequence 41, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-26
```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed protein
US-10-004-381-41

Query Match          100.0%; Score 208; DB 13; Length 426;
Best Local Similarity 100.0%; Pred. No. 3.6e-19;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 378 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 415

RESULT 9
US-10-004-381-38
; Sequence 38, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Designed protein
US-10-004-381-38

Query Match          100.0%; Score 208; DB 13; Length 479;
Best Local Similarity 100.0%; Pred. No. 4.1e-19;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 378 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 415

RESULT 10
US-10-004-381-26
; Sequence 26, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: WILSON, JACK W.
; APPLICANT: SZOSTAK, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-26
```

Query Match 92.8%; Score 193; DB 13; Length 47;
Best Local Similarity 94.7%; Pred. No. 3.3e-18;
Matches 36; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38

RESULT 11
US-10-004-381-27

; Sequence 27, Application US/10004381
; Publication No. US20020155578A1

; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.

; APPLICANT: KEEFE, ANTHONY D.

; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002

; CURRENT APPLICATION NUMBER: US/10/004,381

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 27

; LENGTH: 39

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: selected peptide

US-10-004-381-27

Query Match 71.6%; Score 149; DB 13; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 GHVVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 3 GHVVEGLAGLEQLRRLRLEHHPOGOREP 30

RESULT 12
US-10-004-381-28

; Sequence 28, Application US/10004381
; Publication No. US20020155578A1

; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.

; APPLICANT: KEEFE, ANTHONY D.

; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002

; CURRENT APPLICATION NUMBER: US/10/004,381

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 28

; LENGTH: 89

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: selected peptide

US-10-004-381-28

Query Match 61.1%; Score 127; DB 13; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.9e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 EGLAGLEQLRRLRLEHHPOGOREP 38
Db 3 EGLAGLEQLRRLRLEHHPOGOREP 26

RESULT 13
US-10-004-381-29

; Sequence 29, Application US/10004381
; Publication No. US20020155578A1

; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.

; APPLICANT: KEEFE, ANTHONY D.

; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002

; CURRENT APPLICATION NUMBER: US/10/004,381

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 29

; LENGTH: 29

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: selected peptide

US-10-004-381-29

Query Match 49.0%; Score 102; DB 13; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ELEQLRRLRLEHHPOGOREP 38
Db 2 ELEQLRRLRLEHHPOGOREP 20

RESULT 14
US-10-004-381-11

; Sequence 11, Application US/10004381
; Publication No. US20020155578A1

; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.

; APPLICANT: KEEFE, ANTHONY D.

; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002

; CURRENT APPLICATION NUMBER: US/10/004,381

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 102

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: selected peptide

US-10-004-381-11

Query Match 32.2%; Score 67; DB 13; Length 102;
Best Local Similarity 45.0%; Pred. No. 0.42;
Matches 18; Conservative 2; Mismatches 6; Indels 14; Gaps 3;

Qy 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 28

Fri Jul 16 08:31:13 2004

```
RESULT 15
US-10-156-761-13097
; Sequence 13097, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ. ID NOS: 15109
; SEQ. ID NO 13097
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13097

Query Match      32.0%; Score 66.5; DB 14; Length 739;
Best Local Similarity 32.6%; Pred. No. 4.1;
Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;

QY      1 MDEKTTGWRG-----GHVVEGLAGELBQLRLRLEHHP 32
DB      341 LEDBATGWRGPEPTGMAGPRLARVWNGSENRIEDLRALRLEHHP 383

Search completed: July 15, 2004, 20:45:04
Job time : 36.7647 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:29:44 ; Search time 42.9176 Seconds
(without alignments)
250.172 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208
Sequence: 1 MDEKTTGMRGHHVGEIAGLELDGLRRLRHHHPCQGREP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208	100.0	38	5	ABG67087 Streptavi
2	208	100.0	38	6	ABUS7609 Streptavi
3	208	100.0	47	5	ABG67079 Streptavi
4	208	100.0	47	6	ABUS7601 Synthetic
5	208	100.0	61	5	ABG67078 Streptavi
6	208	100.0	61	6	ABUS7600 Synthetic
7	208	100.0	75	5	ABG67077 Streptavi
8	208	100.0	75	6	ABUS7599 Synthetic
9	208	100.0	89	5	ABG67076 Streptavi
10	208	100.0	89	6	ABUS7598 Synthetic
11	208	100.0	101	5	ABG67075 Streptavi
12	208	100.0	101	6	ABUS7597 Streptavi
13	208	100.0	101	6	ABUS7597 Streptavi
14	208	100.0	101	6	ABUS7595 Streptavi
15	208	100.0	426	5	ABG67091 Maltose-b
16	208	100.0	426	6	ABUS7601 Streptavi
17	208	100.0	479	5	ABG67089 Maltose-b
18	208	100.0	479	6	ABUS7611 Streptavi
19	193	92.8	47	5	ABG67080 Streptavi
20	193	92.8	47	6	ABUS7602 Streptavi
21	149	71.6	39	5	ABG67081 Streptavi
22	149	71.6	39	6	ABUS7603 Synthetic
23	127	61.1	89	5	ABG67082 Streptavi
24	127	61.1	89	6	ABUS7604 Synthetic
25	102	49.0	29	5	ABG67083 Streptavi

26	102	49.0	29	6	ABUS7605 Synthetic
27	67	32.2	102	5	ABG67065 Streptavi
28	67	32.2	102	6	ABUS7587 Streptavi
29	66	31.7	101	5	ABG67066 Streptavi
30	66	31.7	101	6	ABUS7588 Streptavi
31	62	29.8	164	4	ABG63619 Human gas
32	60	28.8	741	4	ABG25060 Novel hum
33	59	28.4	139	6	ADA55387 Human pro
34	59	28.4	140	2	AAW40203 PRAP homo
35	59	28.4	342	3	ABU15903 E. coli p
36	59	28.4	342	6	ABU14680 Protein e
37	59	28.4	538	4	ABG25106 Novel hum
38	58.5	28.1	145	5	ABP09375 Human ORF
39	58	27.9	288	6	ABR96152 Human NOV
40	58	27.9	306	4	AAZ27230 Human EXM
41	58	27.9	314	6	ABR96149 Human NOV
42	58	27.9	318	6	ABR96159 Human NOV
43	58	27.9	318	6	ABR96161 Human NOV
44	58	27.9	318	6	ABR96160 Human NOV
45	58	27.9	318	6	ABR96162 Human NOV

ALIGNMENTS

RESULT 1	ABG67087	standard, peptide, 38 AA.
XX	AC	ABG67087;
XX	XX	
XX	XX	24-SEP-2002 (first entry)
DE	XX	Streptavidin-binding peptide and affinity tag.
XX	XX	Streptavidin-binding peptide.
KW	XX	Synthetic.
OS	XX	WO200238580-A1.
PN	XX	16-MAY-2002.
PD	XX	
XX	XX	31-OCT-2000; 2000WO-US041717.
PF	XX	31-OCT-2000; 2000WO-US041717.
XX	XX	
PR	XX	31-OCT-2000; 2000WO-US041717.
XX	XX	(GENO) GEN HOSPITAL CORP.
PA	XX	Szostak JM, Wilson DS, Keefe AD;
PI	XX	WPI, 2002-500155/53.
DR	XX	
XX	XX	Novel peptide with high affinity for streptavidin, is expressed as part
PT	XX	of fusion protein to facilitate detection, quantitation and purification
PT	XX	of desired protein.
PS	XX	Claim 13; Fig 7A; 56pp; English.
XX	XX	
CC	XX	The invention describes a peptide (I) which binds streptavidin with a
CC	XX	dissociation constant less than 10 mM or 23 nM, where the amino acid
CC	XX	sequence of (I) does not contain an HQQ, HPM, HPN or HQP motif, is not
CC	XX	disulfide bonded or cyclised. A fusion protein comprising a protein of
CC	XX	interest covalently linked to (I) is useful for purifying a desired
CC	XX	protein from a sample by contacting the sample with streptavidin under
CC	XX	conditions that allow complex formation between the fusion protein and
CC	XX	the streptavidin, isolating the complex and recovering the fusion
CC	XX	protein, and thus purifying the desired protein from the sample. The
CC	XX	invention describes a method for producing a streptavidin-binding fusion
CC	XX	protein. This sequence represents a streptavidin-binding peptide used in
CC	XX	the creation of streptavidin-binding DNA-tagged peptide described in the
XX	XX	invention

XX Synthetic peptide SB19 C-terminal deletion mutant, C4.
 XX Streptavidin-binding peptide; mRNA display; peptide library;
 XX fusion protein; mutant; mutein; SB19.
 XX Synthetic.
 OS US2002155578-A1.
 PN 24-OCT-2002.
 PD 31-OCT-2001; 2001US-00004381.
 PF 31-OCT-2000; 2000US-0244541P.
 PR 31-OCT-2000; 2000US-0244541P.
 XX (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX Szostak JW, Wilson DS, Keefe AD;
 PI WPI; 2003-182639/18.
 DR Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX Claim 14; Fig 5; 36pp; English.
 PS The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQQ, HPW, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 CC
 XX Sequence 47 AA;
 SQ
 Query Match 100.0%; Score 208; DB 6; Length 47;
 Best Local Similarity 100.0%; Pred. No. 5.2e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

KW Streptavidin-binding peptide; mutant; mutein.
 XX Escherichia coli.
 OS WO200238580-A1.
 PN 16-MAY-2002.
 PD 31-OCT-2000; 2000WO-US041717.
 PF 31-OCT-2000; 2000WO-US041717.
 PR 31-OCT-2000; 2000WO-US041717.
 XX (GEHO) GEN HOSPITAL CORP.
 PA Szostak JW, Wilson DS, Keefe AD;
 PI WPI; 2002-500155/53.
 DR Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX Claim 9; Fig 5; 56pp; English.
 PS The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQQ, HPW, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 CC
 XX Sequence 61 AA;
 SQ
 Query Match 100.0%; Score 208; DB 5; Length 61;
 Best Local Similarity 100.0%; Pred. No. 7e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PA (WILS/) WILSON D. S.
 XX (KEEFE/) KEEFE A. D.
 PI Szostak JW, Wilson DS, Keefe AD;
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.
 CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulphide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPM, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX
 SQ Sequence 61 AA;
 Query Match 100.0%; Score 208; DB 6; Length 61;
 Best Local Similarity 100.0%; Pred. No. 7e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDEKTTGWRGHHVVEGLAGLEQLRARLEHHHPQGOREP 38
 DB 1 MDEKTTGWRGHHVVEGLAGLEQLRARLEHHHPQGOREP 38
 RESULT 7
 ABG67077 standard; peptide; 75 AA.
 XX
 AC ABG67077;
 DT 24-SEP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #3.
 XX
 KM Streptavidin-binding peptide; mutant; mutein.
 XX
 OS Escherichia coli.
 XX
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX
 PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2002-500155/53.

XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 PS Claim 9; Fig 5; 56pp; English.
 CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 nM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HPO, HPM, HPN or HQP motif, is not
 CC disulphide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX
 SQ Sequence 75 AA;
 Query Match 100.0%; Score 208; DB 5; Length 75;
 Best Local Similarity 100.0%; Pred. No. 8.8e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDEKTTGWRGHHVVEGLAGLEQLRARLEHHHPQGOREP 38
 DB 1 MDEKTTGWRGHHVVEGLAGLEQLRARLEHHHPQGOREP 38
 RESULT 8
 ABU57599 standard; peptide; 75 AA.
 XX
 AC ABU57599;
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C2.
 XX
 KM Streptavidin-binding peptide; mRNA display; peptide library;
 XX
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 XX
 PA (WILS/) WILSON D. S.
 XX
 PA (KEEFE/) KEEFE A. D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.
 CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulphide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain

CC an HPO, HPM, or HOP motif. The peptides are generated by the RNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations
SQ Sequence 75 AA;

Query Match 100.0%; Score 208; DB 6; Length 75;
Best Local Similarity 100.0%; Pred. No. 8.8e-22;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDEKTTGMRGHHVEGLAGLEQLRLRLRLEHHPQGQREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRLRLRLEHHPQGQREP 38

RESULT 9
ABG67076 standard; peptide; 89 AA.

AC ABG67076;
XX
XX 24-SRP-2002 (first entry)
XX
XX Streptavidin-binding peptide SB19 truncation mutant #2.
DE Streptavidin-binding peptide; mutant; mutein.
XX
XX Escherichia coli.
OS
XX
XX MO200238580-A1.
PN
XX
XX 16-MAY-2002.
PD
XX
XX 31-OCT-2000; 2000MO-US041717.
PF
XX
XX 31-OCT-2000; 2000MO-US041717.
PR
XX
XX (GEHO) GEN HOSPITAL CORP.
PA
XX
XX Szostak JW, Wilson DS, Keefe AD;
PI
XX
XX WPI; 2002-500155/53.
DR
XX
XX Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.
XX
XX
XX Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPO, HPM, HPM or HOP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion

CC protein. This sequence represents a truncation mutant of the streptavidin
CC -binding peptide SB-19 isolated from Escherichia coli using a method
CC described in the specification
XX
XX
SQ Sequence 89 AA;

Query Match 100.0%; Score 208; DB 5; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDEKTTGMRGHHVEGLAGLEQLRLRLRLEHHPQGQREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRLRLRLEHHPQGQREP 38

RESULT 10
ABU57598 standard; peptide; 89 AA.

AC ABU57598;
XX
XX 09-APR-2003 (first entry)
XX
XX Synthetic peptide SB19 C-terminal deletion mutant, C1.
DE Synthetic peptide SB19 C-terminal deletion mutant, C1.
XX
XX Streptavidin-binding peptide; mRNA display; peptide library;
KM fusion protein; mutant; mutein; SB19.
XX
XX Synthetic.
OS
XX
XX US200215578-A1.
PN
XX
XX 24-OCT-2002.
PD
XX
XX 31-OCT-2001; 2001US-00004381.
PF
XX
XX 31-OCT-2000; 2000US-0244541P.
PR
XX
XX (SZOS/) SZOSTAK J W.
PA (WILS/) WILSON D S.
PA (KEEF/) KEEFE A D.
XX
XX
XX Szostak JW, Wilson DS, Keefe AD;
PI
XX
XX WPI; 2003-182639/18.
DR
XX
XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.
XX
XX
XX Claim 14; Fig 5; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulfide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPM, HPM or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations

XX
SQ Sequence 89 AA;

Query Match 100.0%; Score 208; DB 6; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 11

ABG67075
ID ABG67075 standard; peptide; 101 AA.

XX
AC ABG67075;

XX
DT 24-SEP-2002 (first entry)

XX
DE Streptavidin-binding peptide SB19 truncation mutant #1.

XX
KM Streptavidin-binding peptide; mutant; mutein.

XX
OS Escherichia coli.

XX
OS Synthetic.

XX
PN WO200238580-A1.

XX
PD 16-MAY-2002.

XX
PF 31-OCT-2000; 2000WO-US041717.

XX
PR 31-OCT-2000; 2000WO-US041717.

XX
PA (GENO) GEN HOSPITAL CORP.

XX
PI Szostak JW, Wilson DS, Keefe AD;

XX
DR MPI; 2002-500155/53.

XX
PT Novel peptide with high affinity for streptavidin, is expressed as part
of fusion protein to facilitate detection, quantitation and purification
of desired protein.

XX
PS Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
disassociation constant less than 10 mM or 23 nM, where the amino acid
sequence of (I) does not contain an HPQ, HPM, HPN or HQP motif, is not
disulfide bonded or cyclised. A fusion protein comprising a protein of
interest covalently linked to (I) is useful for purifying a desired
protein from a sample by contacting the sample with streptavidin under
conditions that allow complex formation between the fusion protein and
streptavidin, isolating the complex and recovering the fusion
protein, and thus purifying the desired protein from the sample. The
invention describes a method for producing a streptavidin-binding fusion
protein. This sequence represents a truncation mutant of the streptavidin
-binding peptide SB-19 isolated from Escherichia coli using a method
described in the specification

XX
SQ Sequence 101 AA;

Query Match 100.0%; Score 208; DB 5; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 12

ABG67073
ID ABG67073 standard; peptide; 101 AA.

XX
AC ABG67073;

XX
DT 24-SEP-2002 (first entry)

XX
DE Streptavidin-binding peptide SB19.

XX
KM Streptavidin-binding peptide.

XX
OS Escherichia coli.

XX
PN WO200238580-A1.

XX
PD 16-MAY-2002.

XX
PF 31-OCT-2000; 2000WO-US041717.

XX
PR 31-OCT-2000; 2000WO-US041717.

XX
PA (GENO) GEN HOSPITAL CORP.

XX
PI Szostak JW, Wilson DS, Keefe AD;

XX
DR MPI; 2002-500155/53.

XX
PT Novel peptide with high affinity for streptavidin, is expressed as part
of fusion protein to facilitate detection, quantitation and purification
of desired protein.

XX
PS Claim 9; Fig 3; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
disassociation constant less than 10 mM or 23 nM, where the amino acid
sequence of (I) does not contain an HPQ, HPM, HPN or HQP motif, is not
disulfide bonded or cyclised. A fusion protein comprising a protein of
interest covalently linked to (I) is useful for purifying a desired
protein from a sample by contacting the sample with streptavidin under
conditions that allow complex formation between the fusion protein and
streptavidin, isolating the complex and recovering the fusion
protein, and thus purifying the desired protein from the sample. The
invention describes a method for producing a streptavidin-binding fusion
protein. This sequence represents a streptavidin-binding peptide isolated
from Escherichia coli using a method described in the specification

XX
SQ Sequence 101 AA;

Query Match 100.0%; Score 208; DB 5; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 13

ABU57597
ID ABU57597 standard; peptide; 101 AA.

XX
AC ABU57597;

XX
DT 09-APR-2003 (first entry)

XX
DE Streptavidin-binding synthetic peptide SB19 full length peptide.

XX
KM Streptavidin-binding peptide; mRNA display; peptide library;

XX
OS Synthetic.

XX
PN US2002155578-A1.

XX 24-OCT-2002.
PD
XX 31-OCT-2001; 2001US-00004381.
XX
PR 31-OCT-2000; 2000US-0244541P.
XX
XX (SZOS/) SZOSTAK J W.
PA (WILS/) WILSON D S.
PA (KEEF/) KEEFE A D.
PI Szostak JW, Wilson DS, Keefe AD;
PI WPI; 2003-182639/18.
DR
XX
XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.
XX
XX Claim 14; Fig 5; 36pp; English.
PS
XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulphide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPM, HPN or HQP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a streptavidin binding peptide of the invention
SQ
XX Sequence 101 AA;
SQ
Query Match 100.0%; Score 208; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDEKTTGMRGSHVVEGLAGELBQLRLARLEHHPQGOREP 38
DB 1 MDEKTTGMRGSHVVEGLAGELBQLRLARLEHHPQGOREP 38
RESULT 14
ABU57595
ID ABU57595 standard; peptide; 101 AA.
XX
XX ABU57595;
AC
XX 09-APR-2003 (first entry)
DT
XX Streptavidin-binding synthetic peptide SB19.
XX
XX Streptavidin-binding peptide; mRNA display; peptide library;
KM fusion protein.
XX
XX Synthetic.
OS
XX US2002155578-A1.
PN
XX 24-OCT-2002.
PD
XX 31-OCT-2001; 2001US-00004381.
XX

PR 31-OCT-2000; 2000US-0244541P.
XX
XX (SZOS/) SZOSTAK J W.
PA (WILS/) WILSON D S.
PA (KEEF/) KEEFE A D.
PI Szostak JW, Wilson DS, Keefe AD;
PI WPI; 2003-182639/18.
DR
XX
XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.
XX
XX Claim 14; Fig 3; 36pp; English.
PS
XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulphide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPM, HPN or HQP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a streptavidin binding peptide of the invention
SQ
XX Sequence 101 AA;
SQ
Query Match 100.0%; Score 208; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDEKTTGMRGSHVVEGLAGELBQLRLARLEHHPQGOREP 38
DB 1 MDEKTTGMRGSHVVEGLAGELBQLRLARLEHHPQGOREP 38
RESULT 15
ABG67091
ID ABG67091 standard; protein; 426 AA.
XX
XX ABG67091;
AC
XX 24-SEP-2002 (first entry)
DT
XX Maltose-binding protein/Streptavidin-binding peptide fusion protein #2.
XX
XX Streptavidin-binding peptide; maltose-binding protein; hexanistidine tag.
KM Escherichia coli.
XX
XX WO200238580-A1.
PN
XX 16-MAY-2002.
PD
XX 31-OCT-2000; 2000WO-US041717.
XX
XX 31-OCT-2000; 2000WO-US041717.
PR
XX (GENO) GEN HOSPITAL CORP.
XX
XX Szostak JW, Wilson DS, Keefe AD;
XX

DR WPI; 2002-500155/53.
DR N-PSDB; ABK95732.
XX

PT Novel peptide with high affinity for streptavidin, is expressed as part
of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

XX
PS Disclosure; Fig 9B; 56pp; English.
XX

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPQ, HPN, HPN or HQP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a fusion protein comprising maltose-
CC binding protein, a streptavidin-binding peptide, and a hexahistidine tag
XX

XX
SQ Sequence 426 AA;

Query Match 100.0%; Score 208; DB 5; Length 426;

Best Local Similarity 100.0%; Pred. No. 6.2e-21;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEBQLRRLHHHPQGQREP 38

DB 378 MDEKTTGMRGSHVVEGLAGLEBQLRRLHHHPQGQREP 415

Search completed: July 15, 2004, 20:34:02
Job time : 42.9176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd

Run on: July 15, 2004, 20:31:20 ; Search time 12.7176 Seconds
(without alignments)
355.490 Million cell updates/sec

Sequence: 1 MDEKTTGWRGGHVVEGLAGE.....LEHHPQGQREPMMSGCKLG 47

Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 283366

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
1: pir78:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	73.5	28.5	554	2	E87275 conserved hypothec
2	65.5	25.4	528	2	C87663 conserved hypothec
3	61	23.6	429	2	F84015 maltose/maltodextr
4	59.5	23.1	342	2	E64905 sugar tr
5	59.5	23.1	342	2	A90894 probable transport
6	59.5	23.1	342	2	H85723 probable transport
7	59	22.9	1974	2	T30010 hypothetical prote
8	58.5	22.7	302	2	B82189 ROK family protein
9	57	22.1	474	1	E24723 tripc protein - Cor
10	57	22.1	496	2	H70668 probable polyketid
11	57	22.1	663	2	D83545 probable helicase
12	56.5	21.9	215	2	A11071 phosphoglycerate p
13	56.5	21.9	1560	2	E15461 SMCX protein - hu
14	56	21.7	361	2	S43218 hypothethical prote
15	56	21.7	328	2	T36025 conserved hypothet
16	55.5	21.5	188	2	E87711 heat shock protein
17	55.5	21.5	288	1	E72671 hypothethical prote
18	55.5	21.5	392	2	F84650 probable protein F
19	35	21.3	324	2	A81364 protein gp45 [Bact
20	35	21.3	377	2	A95265 probable dehydrata
21	54	20.9	147	2	F83505 probable peptid
22	54	20.9	321	2	H87683 transcription regu
23	54	20.9	1156	2	E69444 chromosome segrega
24	53.5	20.7	134	2	AH3177 conserved hypothec
25	53.5	20.7	352	2	H71294 probable o-sialogly
26	53.5	20.7	404	2	S68409 potassium channel
27	53.5	20.7	478	2	C86274 hypothethical prote
28	53	20.5	210	2	E70539 hypothethical prote
29	53	20.5	224	2	D83570 probable nucleotid

30	53	20.5	261	2	C75568	hypochemical prote
31	53	20.5	348	1	S75456	protein plex - syn
32	53	20.5	358	2	C72643	probable 3-denydro
33	53	20.5	528	2	C85056	probable DNA-bindi
34	53	20.5	589	2	S06954	intermediate filam
35	53	20.5	947	2	B86362	hypochemical prote
36	52.5	20.3	184	2	A13507	heat shock protein
37	52.5	20.3	744	2	A82822	NADH-ubiquinone ox
38	52	20.2	113	2	TJQ0810	transcription repr
39	52	20.2	3561	2	T22478	hypochemical prote
40	52	20.2	3761	2	T34734	hypochemical prote
41	52	20.2	378	2	A25399	homeotic protein A
42	52	20.2	450	2	T35496	probable 2-denydro
43	52	20.2	494	2	I52658	neurofilament-66
44	52	20.2	504	2	I53868	alpha-internexin
45	52	20.2	5051	2	A41023	alpha-internexin

ALIGNMENTS

RESULT 1
E87375
conserved hypothetical protein CCI017 [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87375
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Lamb, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-554 <SNO>
A:Cross-References: GB:AE005673; NID:G13422309; PIDN:AAKZ3001.1; GSFDB:GN00148
C:Genetics:
C:Gene: CCI017

Query Match	28.5%	Score 73.5;	DB 2;	Length 554;
Best Local Similarity	35.4%	Pred. No. 0.27;		
Matches 17; Conservative	6;	Mismatches 20;	Indels 5;	Gaps 1

Qy 5 TTGWRGGHVE-----GLAGELQLRLARLEHHPOGOREPMMSGGCKLG 47

Db 45 STEWSAGYFDVNYTFGYYGELNPLRCRLPLLTVGHAAPKIENACELG 92

RESULT 2

conserved hyalothelial protein CC3341 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87663
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidrich,
B.; Laub, M.T.; Desoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.
N., J.; Ermolenko, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; *Proc. Natl. Acad. Sci. U.S.A.* 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249, MUID:21173698, PMID:11259647
A:Accession: C87663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <STO>
A:Cross-references: GB:AE005673, NID:g13425041, PIDN:AAK25303.1, GSPDB:GN001484
C:Genetics:
C:Gene: CC3341

Query Match	25.4%	Score 65.5;	DB 2;	length 528;
Best Local Similarity	37.8%	Pred. No. 2.6;		
Matches 17; Conservative	5;	Mismatches 18;	Indels 5;	Gaps 1;

A:Molecule type: DNA
A:Residues: 1-1974 <DUZ>
A:Cross-references: EMBL:U50309; PIDD:AAAB37057.1; GSPDB:GN00023; CESP:F58G4.1
A:Experimental source: strain Bristol N2; clone F58G4
C:Genetics:
A:Gene: CESP:F58G4.1
A:Map position: 5
A:Introns: 18/3; 164/1; 229/1; 264/1; 378/1; 440/2; 525/3; 1177/2; 1633/3; 1863/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:84-776/Domain: myosin motor domain homology <MMO>

Query Match 22.9%; Score 59; DB 2; Length 1974;
Best Local Similarity 35.7%; Pred. No. 67;
Matches 15; Conservative 7; Mismatches 18; Indels 2; Gaps 1;

QY 8 MRGHHVEGL--AGELEQLRAREHHPQGRPMMSGCCGCG 47
DB 1381 WRKFEGBGVSRABELETTRKLTHTVQEMQOLENANOKIG 1422

RESULT 8

R0K family protein VC1532 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
B82189
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82189
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.D.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; PMID:20406833; PMID:10952301
A:Accession: B82189
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <HEI>
A:Cross-references: GB:AE004231; GB:AE003852; NID:99656027; PIDD:AAF94686.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1532
A:Map position: 1
C:Superfamily: conserved hypothetical protein HI0182; glucose kinase homology

Query Match 22.7%; Score 58.5; DB 2; Length 302;
Best Local Similarity 32.7%; Pred. No. 11;
Matches 16; Conservative 10; Mismatches 12; Indels 11; Gaps 3;

QY 6 TGNRGHVVVEG-----LAGELQLRARE--HHPOGGRPMMSGCG 44
DB 132 TGGCGGLVVEGKVFSGRNHVAGETGMRPLIDMPH-LGEKAPLJAGCGC 179

RESULT 9

E24723
trpC protein - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: E24723
R:Matsumi, K.; Sano, K.; Ohtsubo, E.
Nucleic Acids Res. 14, 10113-10114, 1986
A:Title: Complete nucleotide and deduced amino acid sequences of the Brevibacterium lact
A:Reference number: A93606; PMID:87117512; PMID:3808947
A:Contents: B. lactofermentum
A:Accession: E24723
A:Molecule type: DNA
A:Residues: 1-474 <MAT>
A:Cross-references: GB:X04960; NID:g39591; PIDD:CAA28626.1; PID:g39596
C:Genetics:
A:Gene: trpC
C:Superfamily: trpC-trpF bifunctional enzyme; trpC homology; trpF homology
F:9-258/Domain: trpC homology <TRC>
F:264-460/Domain: trpF homology <TRF>

Query Match 22.1%; Score 57; DB 1; Length 474;
Best Local Similarity 55.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 13 VVEGLAGELEQLRAREH 30
DB 13 IVEGRGHLEETRRARIAH 30

RESULT 10

H70668
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70668; S73074
R:Cole, S.T.; Broesch, K.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: H70668
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-496 <COL>
A:Cross-references: GB:Z83858; GB:AL123456; NID:g3261675; PIDD:CA806102.1; PID:g1781166
A:Experimental source: strain H37RV
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A:Description: Mycobacterium tuberculosis cosmid tbc2.
A:Reference number: S73053
A:Accession: S73074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273; 275-496 <SMI>
A:Cross-references: EMBL:U00024; NID:g560506; PIDD:AAA50929.1; PID:g560508
C:Genetics:
A:Gene: pks15
A:Start codon: GTG
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot
F:86-467/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

Query Match 22.1%; Score 57; DB 2; Length 496;
Best Local Similarity 36.1%; Pred. No. 28;
Matches 13; Conservative 9; Mismatches 12; Indels 2; Gaps 1;

QY 12 HVEGLAGELEQLRAREHHPQGRPE--MMSGCG 45
DB 22 HYLKVAVLEDETRRLAREYEQRAPEPAVVGIGR 57

RESULT 11

D83545
probable helicase PA0799 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83545
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; PMID:20437337; PMID:10984043
A:Accession: D83545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-663 <STO>
A:Cross-references: GB:AE004515; GB:AE004091; NID:g9946687; PIDD:AA604188.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0799

Query Match 22.1%; Score 57; DB 2; Length 663;
Best Local Similarity 36.1%; Pred. No. 38;
Matches 13; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVEGLAGELEQLRARLEHHPOGOR 36
DB 154 LDLEPLAMEGGDNLRFDFARLRFSQARPRATPPQGLR 189

RESULT 12

A11074 phosphoglycerate mutase (EC 5.4.2.1) - *Salmonella enterica* subsp. *enterica* serovar Typhi

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: this species has also been called *Salmonella typhi*

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Oct-2003
C:Accession: A11074

R:Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
A:Reference number: AB0502; NCBI:21534947; PMID:11677608

A:Accession: A11074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-215 <PAR>

A:Cross-references: GB:AL513382; PDB:CAD03416.1; PID:g16505684; GSPDB:GN00176

C:Genetics: gpmB

C:Superfamily: cofactor-dependent phosphoglycerate mutase; phosphoglycerate mutase homol

C:Keyword: intramolecular transferase; isomerase

Query Match 21.9%; Score 56.5; DB 2; Length 215;
Best Local Similarity 26.7%; Pred. No. 14;
Matches 16; Conservative 8; Mismatches 23; Indels 13; Gaps 1;

QY 1 MDEKTTGMR-----GHHVEGLAGELEQLRARLEHHPOGOREPMMGGCKLG 47
DB 96 LTRBEGMRQVLNVGTQDGRIRPGESMGLSDRVNAAALASCELPQGRPLVSHGIALG 155

RESULT 13

154361

SMCX protein - human

N:Alternate names: escapes X-chromosome inactivation

C:Species: *Homo sapiens* (man)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Mar-2000

C:Accession: I54361; I38235

R:Wu, J.; Ellison, J.; Salido, E.; Yen, P.; Mohandas, T.; Shapiro, L.J.

Hum. Mol. Genet. 3, 153-160, 1994

A:Title: Isolation and characterization of XE169, a novel human gene that escapes X-inactivation

A:Reference number: I54361; NCBI:94214434; PMID:8162017

A:Accession: I54361

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1560 <ML>

A:Cross-references: GB:U25270; NID:g457136; PID:g457137

R:Agulnik, A.I.; Mitchell, M.J.; Mattei, M.G.; Borsani, G.; Avner, P.A.; Lerner, J.L.; E

Hum. Mol. Genet. 3, 879-884, 1994

A:Title: A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation

A:Accession: I38235

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 280-341; Y, 343-344 <AGU>

A:Cross-references: EMBL:Z29650; NID:g472994; PID:CAA82758.1; PID:g472995

C:Genetics: gDB:SMCX; XE169; DXS1272E

A:Cross-references: GDB:9865642; OMIM:314690

A:Map position: Xp11.22-Xp11.21

C:Superfamily: human retinoblastoma binding protein 2

Query Match 21.9%; Score 56.5; DB 2; Length 1560;
Best Local Similarity 28.9%; Pred. No. 1,1e+02;
Matches 13; Conservative 7; Mismatches 18; Indels 7; Gaps 1;

QY 1 MDEKTTGMRG-----GHHVEGLAGELEQLRARLEHHPOGOREP 38
DB 1281 LTRALISWGGRRQALASEDVYALIGRLAERQRLQAEPRREBP 1325

RESULT 14

S43218

hypothetical protein YER126c - yeast (*Saccharomyces cerevisiae*)

C:Species: *Saccharomyces cerevisiae*

C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 19-Apr-2002

C:Accession: S43218; S50629

R:Fulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993

A:Reference number: S30812

A:Accession: S43218

A:Molecule type: DNA

A:Residues: 1-261 <ML>

A:Cross-references: GB:U18916; EMBL:L11119; NID:g1384128; PID:AA03224.1; PID:g603365

R:Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: The sequence of *S. cerevisiae* cosmid 9781, 8198, 9115, 9981, and lambda

A:Reference number: S50629

A:Accession: S50629

A:Molecule type: DNA

A:Residues: 1-261 <DIR>

A:Cross-references: EMBL:U18916; NID:g1384128; PID:AA03224.1; PID:g603365; MIPS:YER126

C:Genetics: SGD:KRE32

A:Cross-references: SGD:S0000928

A:Map position: 5R

Query Match 21.7%; Score 56; DB 2; Length 261;
Best Local Similarity 26.7%; Pred. No. 19;
Matches 12; Conservative 10; Mismatches 19; Indels 4; Gaps 1;

QY 3 EKTGMRGHHVEGLAGELEQLRARLEHHPO---GOREPMMGGC 43
DB 40 OKLTGMRGKOPAKRYAEKVSRRKKIKAEOSKVKVSSKPLDTDG 84

RESULT 15

T36025

conserved hypothetical protein SCC54.12c - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-May-2000

C:Accession: T36025

R:Seeger, K.; Harris, D.; Bentley, S.D.; Parhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, March 1999

A:Reference number: Z21581

A:Accession: T36025

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-338 <SER>

A:Cross-references: EMBL:AL035591; PDB:CA838142.1; GSPDB:GN00070; SCOPDB:SCC54.12c

A:Experimental source: strain A3(2)

C:Genetics: gDB:SCC54.12c

C:Superfamily: *Bacillus subtilis* conserved hypothetical protein yycU

Query Match 21.7%; Score 56; DB 2; Length 338;
Best Local Similarity 27.0%; Pred. No. 25;
Matches 17; Conservative 9; Mismatches 11; Indels 26; Gaps 2;

QY 11 GHVVEGLAGELE-----QLRARLEHHPOGOREP-----MMGGC 44
DB 166 GRIVGIAERBLRELFGADLVITSSLVNVEHLRAKMDQAFQDDQEPRLATVMSFGF 225

QY 45 KLQ 47

Fri Jul 16 08:31:15 2004

us-10-004-381-25.rpr

Page 5

Db 226 KYG 228

Search completed: July 15, 2004, 20:36:20
Job time : 13.7176 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:35:51 ; Search time 44.2353 Seconds
(without alignments)
332.096 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDEKTTGMRGHHVGLAGE.....LEHHPGQGRPMWGGCKLG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	258	100.0	47	13	US-10-004-381-25 Sequence 25, Appl
2	243	94.2	47	13	US-10-004-381-26 Sequence 26, Appl
3	241	93.4	61	13	US-10-004-381-24 Sequence 24, Appl
4	234	90.7	75	13	US-10-004-381-23 Sequence 23, Appl
5	227	88.0	89	13	US-10-004-381-22 Sequence 22, Appl
6	214	82.9	101	13	US-10-004-381-21 Sequence 21, Appl
7	211	81.8	101	13	US-10-004-381-19 Sequence 19, Appl
8	208	80.6	38	13	US-10-004-381-35 Sequence 35, Appl
9	208	80.6	426	13	US-10-004-381-41 Sequence 41, Appl
10	208	80.6	479	13	US-10-004-381-38 Sequence 38, Appl
11	199	77.1	29	13	US-10-004-381-27 Sequence 27, Appl
12	152	58.9	29	13	US-10-004-381-29 Sequence 29, Appl
13	133	51.6	89	13	US-10-004-381-28 Sequence 28, Appl
14	94	36.4	44	13	US-10-004-381-34 Sequence 34, Appl
15	70	27.1	102	13	US-10-004-381-11 Sequence 11, Appl

16	68	26.4	101	13	US-10-004-381-12 Sequence 12, Appl
17	66.5	25.8	739	14	US-10-156-761-13097 Sequence 13097, A
18	65	25.2	587	12	US-10-425-114-73019 Sequence 73019, A
19	62.5	24.2	175	16	US-10-437-963-147749 Sequence 147749, A
20	62	24.0	114	12	US-10-425-114-61050 Sequence 61050, A
21	62	24.0	527	12	US-10-425-114-70961 Sequence 70961, A
22	62	24.0	529	12	US-10-425-114-77475 Sequence 37475, A
23	62	24.0	649	12	US-10-425-114-70144 Sequence 70144, A
24	61	23.6	1181	14	US-10-156-761-12448 Sequence 12448, A
25	59.5	23.1	84	11	US-09-864-408A-6610 Sequence 6610, Ap
26	59.5	23.1	342	9	US-09-912-020-260 Sequence 260, App
27	59.5	23.1	1342	12	US-10-282-122A-42604 Sequence 42604, A
28	59	22.9	1391	15	US-10-094-759-2955 Sequence 2955, Ap
29	59	22.9	140	9	US-09-902-180-6 Sequence 6, Appl
30	59	22.9	147	9	US-09-864-761-39307 Sequence 39307, A
31	59	22.9	149	16	US-10-437-963-170766 Sequence 170766, A
32	59	22.9	160	16	US-10-437-963-161643 Sequence 161643, A
33	59	22.9	249	12	US-10-112-944-859 Sequence 859, App
34	59	22.9	249	12	US-10-112-944-860 Sequence 860, App
35	59	22.9	249	12	US-10-112-944-861 Sequence 861, App
36	59	22.9	249	12	US-10-112-944-862 Sequence 862, App
37	59	22.9	869	16	US-10-437-963-156107 Sequence 156107, A
38	59	22.9	1974	15	US-10-369-493-6395 Sequence 6395, Ap
39	58.5	22.7	1336	16	US-10-437-963-115864 Sequence 115864, A
40	58	22.5	115	12	US-10-282-122A-45153 Sequence 45153, A
41	58	22.5	127	16	US-10-437-963-178804 Sequence 178804, A
42	58	22.5	205	12	US-10-425-114-39504 Sequence 39504, A
43	58	22.5	239	12	US-10-425-114-39359 Sequence 39359, A
44	58	22.5	276	12	US-10-425-114-38383 Sequence 38383, A
45	58	22.5	288	12	US-10-112-944-410 Sequence 410, App

ALIGNMENTS

```

RESULT 1
US-10-004-381-25
; Sequence 25, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004.381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-25

Query Match      100.0%; Score 258; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 7e-26;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MDEKTTGMRGHHVGLAGELEQLARLEHHHFGQGRPMWGGCKLG 47
DB      1 MDEKTTGMRGHHVGLAGELEQLARLEHHHFGQGRPMWGGCKLG 47

RESULT 2
US-10-004-381-26
; Sequence 26, Application US/10004381
; Publication No. US20020155578A1

```

```
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 26
/ LENGTH: 47
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-26
```

```
Query Match          94.2%; Score 243; DB 13; Length 47;
Best Local Similarity 95.7%; Pred. No. 6e-24;
Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPMMSGGCKLG 47
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPMMSGGCKLG 47
```

```
RESULT 3
US-10-004-381-24
/ Sequence 24, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 24
/ LENGTH: 61
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-24
```

```
Query Match          93.4%; Score 241; DB 13; Length 61;
Best Local Similarity 77.0%; Pred. No. 1.5e-23;
Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPMMSGGCKL 46
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPLVQVEVDVDEGLVQMSGGCKL 60
```

```
QY 47 G 47
DB 61 G 61
```

```
RESULT 4
US-10-004-381-23
/ Sequence 23, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 23
/ LENGTH: 75
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-23
```

```
Query Match          90.7%; Score 234; DB 13; Length 75;
Best Local Similarity 62.7%; Pred. No. 1.5e-22;
Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPLVQVEVDVDEGLVQDLHGTVAG 38
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPLVQVEVDVDEGLVQDLHGTVAG 60
```

```
QY 39 -----MMSGGCKLG 47
DB 61 LLDPEVKLLTDWPFKKFKNVSMMSGGCKLG 75
```

```
RESULT 5
US-10-004-381-22
/ Sequence 22, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 89
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-22
```

```
Query Match          88.0%; Score 227; DB 13; Length 89;
Best Local Similarity 52.8%; Pred. No. 1.4e-21;
Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPMMSGGCKL 38
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPGQREPLVQVEVDVDEGLVQDLHGTVAG 60
```

```
QY 39 -----MMSGGCKLG 47
DB 61 LLDPEVKLLTDWPFKKFKNVSMMSGGCKLG 89
```

```
RESULT 6
US-10-004-381-21
```



```
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Designed protein
US-10-004-381-38

Query Match          80.6%; Score 208; DB 13; Length 479;
Best Local Similarity 100.0%; Pred. No. 2.6e-18;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHGVHEGLAELEQLRLRLEHHHPQGREP 38
DB 378 MDEKTTGMRGHGVHEGLAELEQLRLRLEHHHPQGREP 415

RESULT 11
US-10-004-381-27
; Sequence 27, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-27

Query Match          77.1%; Score 199; DB 13; Length 39;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GHVVEGLAELEQLRLRLEHHHPQGREPMMSGGCKLG 47
DB 3 GHVVEGLAELEQLRLRLEHHHPQGREPMMSGGCKLG 39

RESULT 12
US-10-004-381-29
; Sequence 29, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
```

```
; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-29

Query Match          58.9%; Score 152; DB 13; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 ELEQLRLRLEHHHPQGREPMMSGGCKLG 47
DB 2 ELEQLRLRLEHHHPQGREPMMSGGCKLG 29

RESULT 13
US-10-004-381-28
; Sequence 28, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-28

Query Match          51.6%; Score 133; DB 13; Length 89;
Best Local Similarity 35.6%; Pred. No. 2e-09;
Matches 31; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY 15 EGLAELEQLRLRLEHHHPQGREPMM----- 40
DB 3 EGLAELEQLRLRLEHHHPQGREPLVQREVDVDEGLVDLHGVAGLDPVEKLLTDMFK 62

QY 41 -----SGGCKLG 47
DB 63 KFKVSKDCKMTFYLEMVDMSGGCKLG 89

RESULT 14
US-10-004-381-34
; Sequence 34, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
```

```

; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed peptide
US-10-004-381-34

```

```

Query Match          36.4%; Score 94; DB 13; Length 44;
Best Local Similarity 58.3%; Pred. No. 0.0001;
Matches 21; Conservative 3; Mismatches 4; Indels 8; Gaps 2;

```

```

QY      18 AGELEQLRAR-----LEHHPOGOREPMMGGCKTG 47
          |||:|||:|||:|||:|||:|||:|||:|||
DB      8 AGPVDAQADARLVQAGALQHHPOGDR--MMSGCKTG 41

```

```

RESULT 15
US-10-004-381-11
; Sequence 11, Application US/10004381
; Publication No. US2002015578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-11

```

```

Query Match          27.1%; Score 70; DB 13; Length 102;
Best Local Similarity 42.9%; Pred. No. 0.33;
Matches 18; Conservative 4; Mismatches 6; Indels 14; Gaps 3;

```

```

QY      1 MDEKTTGW--RGHVVVGLAGLEQLRARLEHHPOGOREPMM 40
          |||:|||:|||:|||:|||:|||:|||:|||
DB      1 MDEK-TTHMRVYHHLADG-----LEQHPQGGRPLV 30

```

Search completed: July 15, 2004, 20:45:03
 Job time : 45.2353 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:32:20 ; Search time 16.0353 Seconds

(without alignments)
151.318 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDKKTGMGSHVEGLAGE.....LEHHPGQREPMMSGCKXG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	24.4	543	3 US-09-413-814-91	Sequence 91, Appl
2	63	24.4	544	3 US-09-413-814-81	Sequence 81, Appl
3	62	24.0	356	4 US-09-252-991A-19549	Sequence 19549, A
4	61	23.6	239	4 US-09-134-000C-5025	Sequence 5025, Ap
5	60	23.3	152	4 US-09-252-991A-18469	Sequence 18469, A
6	60	23.3	432	4 US-09-252-991A-21794	Sequence 21794, A
7	59.5	23.1	543	4 US-09-252-991A-29830	Sequence 29830, A
8	58.5	22.7	197	4 US-09-252-991A-24825	Sequence 24825, A
9	58	22.5	113	2 US-08-248-839C-6	Sequence 6, Appl
10	58	22.5	207	4 US-09-328-352-5708	Sequence 5708, Ap
11	58	22.5	266	4 US-09-252-991A-27885	Sequence 27885, A
12	58	22.5	628	4 US-09-252-991A-22131	Sequence 22131, A
13	57.5	22.3	1065	3 US-09-412-545-2	Sequence 2, Appl
14	57	22.1	296	4 US-09-252-991A-30737	Sequence 30737, A
15	57	22.1	495	4 US-08-311-731A-3	Sequence 3, Appl
16	57	22.1	674	4 US-09-252-991A-18107	Sequence 18107, A
17	57	22.1	904	4 US-09-252-991A-29119	Sequence 29119, A
18	57	22.1	1136	4 US-09-252-991A-31394	Sequence 31394, A
19	56.5	21.9	220	4 US-09-252-991A-24796	Sequence 24796, A
20	56.5	21.9	283	4 US-09-252-991A-20071	Sequence 20071, A
21	56	21.7	468	4 US-09-199-637A-29	Sequence 29, Appl
22	55.5	21.5	1008	4 US-09-252-991A-19329	Sequence 19329, A
23	55.5	21.5	1388	4 US-09-572-191-2	Sequence 2, Appl
24	55.5	21.5	1388	4 US-09-723-262-2	Sequence 2, Appl
25	55.5	21.5	1388	4 US-09-723-219-2	Sequence 2, Appl
26	55	21.3	162	4 US-09-252-991A-20901	Sequence 20901, A
27	55	21.3	184	4 US-09-252-991A-30468	Sequence 30468, A

28	55	21.3	578	4 US-09-252-991A-31318	Sequence 31318, A
29	54.5	21.1	459	4 US-09-134-000C-6639	Sequence 6639, Ap
30	54.5	21.1	478	4 US-09-252-991A-31542	Sequence 31542, A
31	54.5	21.1	602	4 US-09-252-991A-30529	Sequence 30529, A
32	54.5	21.1	1841	2 US-08-804-227C-6	Sequence 6, Appl
33	54.5	21.1	4630	4 US-09-091-609-2	Sequence 2, Appl
34	54.5	21.1	5215	3 US-09-105-537-2	Sequence 2, Appl
35	54	20.9	185	4 US-09-252-991A-19647	Sequence 19647, A
36	54	20.9	499	4 US-09-252-991A-27221	Sequence 27221, A
37	54	20.9	631	4 US-09-252-991A-26444	Sequence 26444, A
38	53.5	20.7	266	4 US-09-252-991A-24870	Sequence 24870, A
39	53.5	20.7	511	4 US-09-252-991A-18418	Sequence 18418, A
40	53.5	20.7	692	4 US-09-252-991A-19069	Sequence 19069, A
41	53	20.5	224	4 US-09-199-637A-175	Sequence 175, Ap
42	53	20.5	258	4 US-09-252-991A-17377	Sequence 17377, A
43	53	20.5	648	4 US-09-252-991A-23682	Sequence 23682, A
44	53	20.5	752	4 US-09-252-991A-17355	Sequence 17355, A
45	52.5	20.3	69	4 US-09-328-352-4132	Sequence 4132, Ap

ALIGNMENTS

```

RESULT 1
US-09-413-814-91
; Sequence 91, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoffe, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 91
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-91

Query Match      24.4% Score 63; DB 3; Length 543;
Best Local Similarity 35.1%; Pred. No. 2.1;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

Db      8 WRGSHVEG-----LAGELEQRLARLEHHPGQR 36
235 WLSGEVLBEGLARVWTKLSGALRRARVPVDEHPARR 271

RESULT 2
US-09-413-814-81
; Sequence 81, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra

```

```
/ APPLICANT: Cino, Paul M
/ APPLICANT: Dougherty, Brian A
/ APPLICANT: Goldberg, Steven L
/ APPLICANT: Hofle, Gerhard
/ APPLICANT: Mueller, Joachim
/ APPLICANT: Reichenbach, Hans
/ TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
/ FILE REFERENCE: PCT/US 99/23535
/ CURRENT APPLICATION NUMBER: US/09/413,814
/ CURRENT FILING DATE: 1999-10-07
/ EARLIER APPLICATION NUMBER: DE 198 46 493.2
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 81
/ LENGTH: 544
/ TYPE: PRT
/ ORGANISM: Sorangium cellulosum
/ US-09-413-814-81

Query Match          24.4%; Score 63; DB 3; Length 544;
Best Local Similarity 35.1%; Pred. No. 2.1;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY      8 WRGCHVEG-----LAGELEQLRLEHHHPOGOR 36
DB      236 WLSGEVLEGLRWTKLSGALRAKAVPVDHPEAGNR 272

RESULT 3
US-09-252-991A-19549
/ Sequence 19549, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 19549
/ LENGTH: 356
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-19549

Query Match          24.0%; Score 62; DB 4; Length 356;
Best Local Similarity 41.0%; Pred. No. 1.8;
Matches 16; Conservative 5; Mismatches 12; Indels 6; Gaps 2;

QY      7 GWRGCHVEGLAGLEQLRLEHHHPO--GOREPMMSG 43
DB      183 GHRGCHAGAGPEPQ----RATBEQHPKAVGQROOQIABG 217

RESULT 4
US-09-134-000C-5025
/ Sequence 5025, Application US/09134000C
/ Patent No. 6617156
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
```

```
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5025
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-134-000C-5025

Query Match          23.6%; Score 61; DB 4; Length 239;
Best Local Similarity 48.5%; Pred. No. 1.5;
Matches 16; Conservative 3; Mismatches 10; Indels 4; Gaps 2;

QY      8 WRGCHVEGLAGLEQLRLEHHHPOGOREPMM 40
DB      34 W-GSEVITGLMASMEQERAKEHE--GRETMM 62

RESULT 5
US-09-252-991A-18469
/ Sequence 18469, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 18469
/ LENGTH: 152
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-18469

Query Match          23.3%; Score 60; DB 4; Length 152;
Best Local Similarity 36.8%; Pred. No. 1.2;
Matches 14; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY      10 GCHVVEGLAGLEQLRLEHHHPOGOREPMMSGCKG 47
DB      74 GGMVVEVAGLVEQYQLRTVDHRAQDRHPLAAGQLG 111

RESULT 6
US-09-252-991A-21794
/ Sequence 21794, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 21794
/ LENGTH: 432
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (405)
/ OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
```

US-09-252-991A-21794

Query Match 23.3%; Score 60; DB 4; Length 432;

Best Local Similarity 52.9%; Pred. No. 4.1;

Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 27 RLEHHPGQREPMMSG 43

Db 143 RLHHHGGRRQPYLAG 159

RESULT 7

US-09-252-991A-29830

; Sequence 29830; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074.788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094.190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29830

; LENGTH: 543

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29830

Query Match 23.1%; Score 59.5; DB 4; Length 543;

Best Local Similarity 42.9%; Pred. No. 6.3;

Matches 15; Conservative 3; Mismatches 14; Indels 3; Gaps 2;

QY 7 GWRGHHVEGLAG--ELEQLARLEHHPG--GOREP 38

Db 183 GGSAGHAFQGLADLVLDQLGVAKRHHRPDMRHP 217

RESULT 8

US-09-252-991A-24825

; Sequence 24825; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074.788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094.190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24825

; LENGTH: 197

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24825

Query Match 22.7%; Score 58.5; DB 4; Length 197;

Best Local Similarity 44.8%; Pred. No. 2.7;

Matches 13; Conservative 2; Mismatches 13; Indels 1; Gaps 1;

QY 8 WRGHHVEGLAGLEQLARL--EHHPOGQ 35

Db 112 WRGGAEPGDPGRRRRRLPLRRHHPOGR 140

RESULT 9

US-08-248-839C-6

; Sequence 6; Application US/08248839C

; Patent No. 5843702

; GENERAL INFORMATION:

; APPLICANT: McConnell, David

; APPLICANT: Devine, Kevin

; APPLICANT: O'Kane, Charles

; TITLE OF INVENTION: A Gene Expression System

; NUMBER OF SEQUENCES: 185

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702ch America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/248.839C

; FILING DATE: 25-MAY-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valela A.

; REGISTRATION NUMBER: 35,127

; REFERENCE/DOCKET NUMBER: 3614.214-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 113 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

US-08-248-839C-6

Query Match 22.5%; Score 58; DB 2; Length 113;

Best Local Similarity 40.0%; Pred. No. 1.7;

Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 9 RGHGHHVEGLAGLEQLARLEHHPGQOREP 38

Db 10 RGRKQDEIAGHIGVSRARYSHYENGSRSP 39

RESULT 10

US-09-328-352-5708

; Sequence 5708; Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328.352

; PRIOR FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5708

; LENGTH: 207

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-5708

Query Match 22.5%; Score 58; DB 4; Length 207;

Best Local Similarity 45.5%; Pred. No. 3.3;

Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 22 EQLRARLEHHPGQOREPMMSG 43

Db 106 QVLSRRFSSHPLMKKEPMATRG 127

RESULT 11

US-09-252-991A-27885
; Sequence 27885, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27885
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27885

Query Match 22.5%; Score 58; DB 4; Length 286;

Best Local Similarity 33.3%; Pred. No. 4.8; Mismatches 23; Indels 0; Gaps 0;

Matches 15; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

Qy 3 EKTGRRGGHVVAGGLAEQLRLAEHHHPOGOREPMMGGCKLG 47

Db 78 EHTHDPGRSRRVAVAGGFVRQQGRLVDDAGNADPLLAGKAVG 122

RESULT 12

US-09-252-991A-22131
; Sequence 22131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22131
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22131

Query Match 22.5%; Score 58; DB 4; Length 628;

Best Local Similarity 44.4%; Pred. No. 12; Mismatches 11; Indels 8; Gaps 1;

Matches 16; Conservative 1; Mismatches 11; Indels 8; Gaps 1;

Qy 9 RGGHVV-----GAGLEQLRLAEHHHPOGOR 36

Db 126 RGGHVVVVGLAVGGGLAVELGAVPRGADHPVGPR 161

RESULT 13

US-09-412-545-2
; Sequence 2, Application US/09412545
; Patent No. 6255095
; GENERAL INFORMATION:
; APPLICANT: Prescott, Stephen M.
; APPLICANT: Ding, Li

APPLICANT: Traer, Elie
; TITLE OF INVENTION: HUMAN DIACYLGLYCEROL KINASE IOTA
; FILE REFERENCE: 1321.2.25
; CURRENT APPLICATION NUMBER: US/09/412,545
; CURRENT FILING DATE: 1999-10-05
; EARLIER APPLICATION NUMBER: 60/103,079
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-412-545-2

Query Match 22.3%; Score 57.5; DB 3; Length 1065;

Best Local Similarity 25.9%; Pred. No. 26; Mismatches 14; Conservative 9; Mismatches 20; Indels 11; Gaps 1;

Matches 14; Conservative 9; Mismatches 20; Indels 11; Gaps 1;

Qy 4 KTTGRRGGH-----VVEGLAEQLRLAEHHHPOGOREPMMGGCKL 46

Db 469 KTLNMGGITDEPVSKILQVYEDGTVVOLDKRWNLHVERNPDLPELEDGVCKL 522

RESULT 14

US-09-252-991A-30737
; Sequence 30737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30737
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30737

Query Match 22.1%; Score 57; DB 4; Length 296;

Best Local Similarity 46.2%; Pred. No. 6.9; Mismatches 12; Indels 0; Gaps 0;

Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 18 AGLEQLRLAEHHHPOGOREPMMGG 43

Db 55 AGVLDLRLVAAEHHRMLRGQFQVGG 80

RESULT 15

US-08-311-731A-3
; Sequence 3, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210


```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
; US-08-311-731A-3

```

```

Query Match      22.1%; Score 57; DB 4; Length 495;
Best Local Similarity 36.1%; Pred. No. 12;
Matches 13; Conservative 9; Mismatches 12; Indels 2; Gaps 1;

QY      12 HYVEGLAGELBOARLEHHNPOGQRRP--MMSGGCK 45
DB      22 HYIKKVAVELEDTRRRLREYEQRAPEVAVVGIGCR 57

```

Search completed: July 15, 2004, 20:37:00
Job time : 17.0353 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:55 ; Search time 29.5059 Seconds

(without alignments)
406.349 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208
Sequence: 1 MDEKTGMGRGHVYBGLRGLRLRLHHPGQREP 38

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66.5	32.0	739	16	Q82BY9 streptomyc
2	65.5	31.5	189	2	Q8KRD7 rhodococcu
3	61	29.3	773	2	033541 endosymbion
4	60	28.8	588	16	Q8DH85 synecococc
5	59	28.4	139	4	Q96MD1 homo sapien
6	59	28.4	342	16	Q8XAY8 escherichia
7	59	28.4	342	16	Q83RD8 shigella fl
8	59	28.4	1083	3	094188 cryptococcu
9	59	28.4	2360	3	094188 cryptococcu
10	58.5	28.1	274	16	Q8P4G4 xanthomonas
11	58	27.9	333	4	Q8IYU4 homo sapien
12	58	27.9	491	2	Q9ZCD9 streptomyc
13	58	27.9	1161	16	Q9JOL0 streptomyc
14	58	27.9	1181	16	Q82DQ5 streptomyc
15	57	27.4	229	16	Q7U554 synecococc
16	57	27.4	489	16	Q8FLJ7 corynebacte

17	57	27.4	663	16	Q91SD9 pseudomonas
18	56.5	27.2	274	2	Q83V70 halomonas s
19	56.5	27.2	377	16	Q9RK62 streptomyc
20	56.5	27.2	1559	4	Q7Z5S5 homo sapien
21	56	26.9	600	2	Q7WUL9 pseudomonas
22	56	26.9	717	11	Q8BMW4 mus musculus
23	56	26.9	976	5	Q9U158 leishmania
24	55.5	26.7	290	16	Q82ZC6 enterococcu
25	55.5	26.7	360	16	Q882M7 pseudomonas
26	55.5	26.7	430	5	Q9W4Q1 diosiphila
27	55.5	26.7	440	16	Q8K692 streptococ
28	55.5	26.7	442	16	Q879H3 streptococ
29	55.5	26.7	1388	4	Q9NS87 streptomyc
30	55	26.4	134	4	Q8WVX5 homo sapien
31	55	26.4	453	16	Q881A2 pseudomonas
32	55	26.4	487	13	Q8JGR1 brachydanio
33	55	26.4	487	13	Q7ZU28 brachydanio
34	55	26.4	605	4	Q96NPF9 homo sapien
35	55	26.4	714	16	Q89JUN0 bradyrhizob
36	55	26.4	1206	10	Q7XNE8 oryza sativ
37	54.5	26.2	536	5	Q8T9S0 aplysia cal
38	54.5	26.2	536	5	Q8T0Y9 aplysia cal
39	54.5	26.2	786	16	Q7V5N6 prochloroco
40	54.5	26.2	835	10	Q81H08 oryza sativ
41	54	26.0	159	16	Q82J34 streptomyc
42	54	26.0	321	16	Q9A2D0 caulobacter
43	54	26.0	329	10	Q8L6S5 oryza sativ
44	54	26.0	469	16	Q8PRC0 xanthomonas
45	54	26.0	674	11	Q60979 mus musculus

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	739 AA.
Q82BY9	Q82BY9		
AC	Q82BY9		
DT	01-JUN-2003 (TREMBLrel_24, Created)		
DT	01-JUN-2003 (TREMBLrel_24, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel_25, Last annotation update)		
DE	Putative regulatory protein.		
GN	SAV5565.		
OS	Streptomyces avermitilis.		
OC	Bacteria; Actinobacteria; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.		
OX	NCBI_Taxid=33903;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=21477403; PubMed=11572948;		
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,		
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,		
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;		
RT	"Genome sequence of an industrial microorganism Streptomyces		
RT	avermitilis: deducing the ability of producing secondary		
RT	metabolites.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=22608306; PubMed=12692562;		
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,		
RA	Sakaki Y., Hattori M., Omura S.;		
RT	"Complete genome sequence and comparative analysis of the industrial		
RT	microorganism Streptomyces avermitilis.";		
RT	Nat. Biotechnol. 21:526-531(2003).		
EMBL	AB005043; BAC73277.1; .		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0003824; F:catalytic activity; IEA.		
DR	InterPro; IPR003594; A:ATPase.		
DR	InterPro; IPR001932; P:PC-1like.		
DR	Pfam; PF02518; HATPase_c; 1.		

```

DR SMART: SM00331; PP2C_SIG; 1.
KM Complete proteome.
SQ SEQUENCE 739 AA; 77478 MW; 63853F93EC7A2B6 CRC64;

Query Match
Best Local Similarity 32.0%; Score 66.5; DB 16; Length 739;
Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;

QY 1 MDEKTTGMRG-----GHVVGGLAGLEQLPARLEHHP 32
DB 341 LBDGATGMRGPETGMAPGPRLAQVWVGSENRLEDLRALERHP 383

RESULT 2
Q8KRD7 PRELIMINARY; PRT; 189 AA.
AC Q8KRD7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transcription repressor KfrA-like protein.
OS Rhodococcus erythropolis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MP50;
RA Trott S., Burger S., Calaminus C., Stolz A.;
RT "Cloning and Heterologous Expression of an Eranthisselective Amidase
from Rhodococcus erythropolis Strain MP50."
RL Appl. Environ. Microbiol. 68:3279-3286(2002).
DR EMBL; AY026386; AKL1725.1; -.
SQ SEQUENCE 189 AA; 20101 MW; 4B8B5D00A093FFFD CRC64;

Query Match
Best Local Similarity 45.5%; Score 65.5; DB 2; Length 189;
Matches 20; Conservative 2; Mismatches 13; Indels 9; Gaps 1;

QY 3 EKTGWRGSHV-----VEGLAGLEQLPARLEHHPGORE 37
DB 104 EQHTAERGDAVTARDRALGEVEGLRVLEQLRGLETARQDARE 147

RESULT 3
Q33541 PRELIMINARY; PRT; 773 AA.
AC Q33541;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Sensor Protein RsaA.
GN RsaA.
OS endosymbiont of Ricketia pachyptila.
OC Bacteria; Proteobacteria; Gammaproteobacteria;
OC sulfur-oxidizing symbionts.
OX NCBI_TaxID=54396;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97438512; PubMed=9293000;
RA Hughes D.S., Felbeck H., Stein J.L.;
RT "A histidine protein kinase homolog from the endosymbiont of the
hydrothermal vent tubeworm Ricketia pachyptila."
RL Appl. Environ. Microbiol. 63:3494-3498(1997).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
KINASES.
DR EMBL; U93704; AAB71131.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```

```

DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0000760; F:sensory perception; IEA.
DR GO; GO:0000160; F:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR003594; AtPbld ATPase.
DR InterPro; IPR004358; Bact_sens_pr_C.
DR InterPro; IPR005467; His_Kinase.
DR InterPro; IPR003661; His_kina_N.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00512; HisKA_1.
DR Pfam; PF00072; response_reg; 1.
DR PRINTS; PR00344; BCTRISENSOR.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00448; REC; 1.
DR SMART; SM00388; HisKA; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 773 AA; 87086 MW; 2C5643B64F373B27 CRC64;

Query Match
Best Local Similarity 43.3%; Score 61; DB 2; Length 773;
Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 8 WRGSHVVEGLAGLEQLPARLEHHPGORE 37
DB 346 WRGPREIELEKLTALAEYHAHHTQORE 375

RESULT 4
Q8DH85 PRELIMINARY; PRT; 588 AA.
AC Q8DH85;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Penicillin-binding protein.
GN TLR2074.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005376; BAC09626.1; -.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009227; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
KW Complete proteome.
SQ SEQUENCE 588 AA; 65048 MW; 9E24A8544A102719 CRC64;

Query Match
Best Local Similarity 28.8%; Score 60; DB 16; Length 588;
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 12 HVEGLAGLEQLPARLEHHP 32
DB 442 HVIDGLYDNGLOQLRLEHHP 462

```

```

RESULT 5
O96MD1 PRELIMINARY; PRT; 139 AA.
AC O96MD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ32575.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SpLeen;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Oca T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamaashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isegai T.;
RA "MEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK057137; BAB71369.1; -.
KW Hypothetical protein.
SQ SEQUENCE 139 AA; 15014 MW; 3A59D86197848F19 CRC64;

Query Match 28.4%; Score 59; DB 4; Length 139;
Best Local Similarity 44.8%; Pred. No. 11;
Matches 13; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Oy 3 EKTGWGCHVEGLAGELBQLRLRL 31
Db 92 ERATGMRLSHEVSGSGQLATSCR-HH 118

RESULT 6
O8XAY8 PRELIMINARY; PRT; 342 AA.
AC O8XAY8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative transport system permease protein.
GN YDEY OR Z2191 OR EGS2121.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Groetbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RL "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005353; AAG56252.1; -.

```

```

DR EMBL; AP002557; BAB35544.1; -.
DR PIR; A90894; A90894.
DR PIR; H85723; H85723.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 36365 MW; 6953A23C07C305A8 CRC64;

Query Match 28.4%; Score 59; DB 16; Length 342;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 8 WRGCHVEGLAGELBQLRLRL 28
Db 135 WTGKWIIEGLPAELKQLSAPL 155

RESULT 7
O83RD8 PRELIMINARY; PRT; 342 AA.
AC O83RD8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative transport system permease protein.
GN YDEY OR SP1584 OR S1710.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RL "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RL "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE015180; AA043171.1; -.
DR EMBL; AE015983; AAP17063.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 36396 MW; E5FE60B2AA54604B CRC64;

Query Match 28.4%; Score 59; DB 16; Length 342;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 8 WRGCHVEGLAGELBQLRLRL 28
Db 135 WTGKWIIEGLPAELKQLSAPL 155

```

```
RESULT 8
ID 094189 PRELIMINARY; PRT; 1083 AA.
AC 094189;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Phosphatidylinositol 3-kinase TOR1 (Fragment).
GN TOR1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxId=40410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B3501;
RX MEDLINE=9262981; PubMed=10330150;
RA Cruz M.C., Cavallo L.M., Gorlach J.M., Cox G., Perfect J.R., Cardenas M.B., Heltman J.;
RT Rapamycin antifungal action is mediated via conserved complexes with FKBP12 and TOR kinase homologs in Cryptococcus neoformans.";
RL Mol. Cell. Biol. 19:4101-4112(1999).
DR EMBL; AF098973; AAD16274.1; -.
DR HSSP; PA2345; LAUF.
DR GO; GO:0004428; F:inositol/phosphatidylinositol kinase activity; IEA.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR000403; P13_P14_kinase.
DR InterPro; IPR008941; TPR-1-like.
DR Pfam; PF02259; FAT; 1.
DR Pfam; PF02260; FATC; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00146; PI3K; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS0290; P13_4_KINASE_3; 1.
KW Kinase.
FT NON_TER
FT NON_TER
SQ SEQUENCE 1083 AA; 123973 MW; 0C48A17D758353A9 CRC64;

Query Match 28.4%; Score 59; DB 3; Length 1083;
Best Local Similarity 35.1%; Pred. No. 1e+02;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVGEGLAEQLRLARLEHHHPOGORE 37
DB 581 LEEASKHYFGDHDIPGMLGVLEPLHEIIVENGPOTLRE 617

RESULT 9
ID 094188 PRELIMINARY; PRT; 2360 AA.
AC 094188;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Phosphatidylinositol 3-kinase TOR1.
GN TOR1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxId=40410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RX MEDLINE=9262981; PubMed=10330150;
RA Cruz M.C., Cavallo L.M., Gorlach J.M., Cox G., Perfect J.R., Cardenas M.B., Heltman J.;
RT Rapamycin antifungal action is mediated via conserved complexes with FKBP12 and TOR kinase homologs in Cryptococcus neoformans.";
RL Mol. Cell. Biol. 19:4101-4112(1999).
```

```
DR EMBL; AF098972; AAD16273.1; -.
DR HSSP; PA2345; LAUF.
DR GO; GO:0004428; F:inositol/phosphatidylinositol kinase activity; IEA.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR000403; P13_P14_kinase.
DR Pfam; PF02259; FAT; 1.
DR Pfam; PF02260; FATC; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00146; PI3K; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS0290; P13_4_KINASE_3; 1.
KW Kinase.
SQ SEQUENCE 2360 AA; 267216 MW; ED4A1059B1AA2B2A CRC64;

Query Match 28.4%; Score 59; DB 3; Length 2360;
Best Local Similarity 35.1%; Pred. No. 2.4e+02;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVGEGLAEQLRLARLEHHHPOGORE 37
DB 1858 LEEASKHYFGDHDIPGMLGVLEPLHEIIVENGPOTLRE 1894

RESULT 10
ID 08P4G4 PRELIMINARY; PRT; 274 AA.
AC 08P4G4;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Lipopolysaccharide core biosynthesis glycosyl transferase DE (Glycosyltransferase).
GN XCC3744 OR WAXE.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxId=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R., Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F., Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., Camarotte G., Camavau J., Cardoso J., Chambergo F., Clapina L.P., Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T., Formighieri E.F., Franco M.C., Greggio C.C., Gruber A., Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F., Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M., Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H., Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R., Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F., Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D., Trindade dos Santos M., Truffi D., Teal S.M., White F.F., Setubal J.C., Kiteajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing host specificities.";
RT Nature 417:459-463(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8004;
RX Lu G., Tang J., He Y., Chen B., Tang D.;
RT "Identification and Cloning of a Novel Gene Involved in EPS Biosynthesis of Xanthomonas campestris pv. campestris.";
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AE012495; AAM43001.1; -.
DR EMBL; AY129625; AAP3588.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001173; Glyco_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase; Complete proteome.
```

SQ SEQUENCE 274 AA; 31016 MW; 28F541FA0102D64B CRC64;
 Query Match 28.1%; Score 58.5; DB 16; Length 274;
 Best Local Similarity 38.2%; Pred. No. 26;
 Matches 13; Conservative 3; Mismatches 15; Indels 3; Gaps 1;
 QY 2 DEKTTGMRGCHVVEGLAGELEQRLARLEHHHP 32
 DB 144 DRRGGGMRGKREIHEAASVDGTVATLRGLDIIHP 177
 RESULT 11
 Q81U04 PRELIMINARY; PRT; 333 AA.
 AC 081U04;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (Nov-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC040438; AAH0438.1; -.
 DR InterPro; IPR001073; ClQ.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008983; TNF_like.
 DR Pfam; PF00386; Clq; 1.
 DR Pfam; PF01381; Collagen; 3.
 DR PRINTS; PR00007; COMPLEMENTC1Q.
 DR SMART; SM00110; ClQ; 1.
 DR PROSITE; PS01113; ClQ; 1.
 DR KMW Hypothetical protein.
 SQ SEQUENCE 333 AA; 34649 MW; E9EBF803A034E307 CRC64;
 Query Match 27.9%; Score 58; DB 4; Length 333;
 Best Local Similarity 40.0%; Pred. No. 37;
 Matches 14; Conservative 6; Mismatches 15; Indels 0; Gaps 0;
 QY 2 DEKTTGMRGCHVVEGLAGELEQRLARLEHHHP 36
 DB 90 DQSGRSGPKRGKGLAGPWEKGLRGETGPGQOK 124
 RESULT 12
 Q9ZGD9 PRELIMINARY; PRT; 491 AA.
 AC 09ZGD9;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Oxygenase homolog.
 GN LANE.
 OS Streptomyces cyanogenus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=80860;
 RX SEQUENCE FROM N.A.
 RC STRAIN=S136;
 RC MEDLINE=99132695; PubMed=9933932;
 RA Westrich L., Domann S., Faust B., Bedford D., Hopwood D.A.,
 RA Bechthold A.;
 RT "Cloning and characterization of a gene cluster from Streptomyces
 cyanogenus S136 probably involved in landomycin biosynthesis.";
 RL FEMS Microbiol. Lett. 170:381-387 (1999).
 DR EMBL; AF080235; AAD13534.1; -.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.

DR GO; GO:0006725; P:aromatic compound metabolism; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000733; P:cytochrome P-450 monooxygenase.
 DR InterPro; IPR002938; Mox_FAD_binding.
 DR InterPro; IPR003025; NAD_BS.
 DR InterPro; IPR003042; Rng_mnoxygenase.
 DR Pfam; PF01494; FAD binding 3; 1.
 DR Pfam; PF01360; Monooxygenase; 1.
 DR PRINTS; PR00420; RINGMONOXENASE.
 SQ SEQUENCE 491 AA; 52062 MW; 096FD4A642275941 CRC64;
 Query Match 27.9%; Score 58; DB 2; Length 491;
 Best Local Similarity 40.5%; Pred. No. 57;
 Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;
 QY 2 DEKTTGMRGCHVVEGLAGE-LEQRLARLEHHHP 36
 DB 293 DSVNLGKMLAAVNGIAGGLDSYHS--ERHPVGR 327
 RESULT 13
 Q910L0 PRELIMINARY; PRT; 1161 AA.
 AC 0910L0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DNA-directed RNA polymerase beta chain.
 GN RPOB OR SCO4654 OR SCD82.26.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=1902;
 RX SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=843436;
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RC MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Lake L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)." ;
 RL Nature 417:141-147 (2002).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 {RNA} (N).

```

CC -1- SUBUNIT. THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
DR EMBL: A939121, CAB7428.1; -.
DR HSSP: Q9KWTJ, IJOM.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0016740; F:transcriptase activity; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR007121; RNA_pol_B.
DR InterPro: IPR007644; RNA_pol_Rpb2_1.
DR InterPro: IPR007642; RNA_pol_Rpb2_2.
DR InterPro: IPR007645; RNA_pol_Rpb2_3.
DR InterPro: IPR007120; RNA_pol_Rpb2_6.
DR InterPro: IPR007641; RNA_pol_Rpb2_7.
DR Pfam: PF04563; RNA_pol_Rpb2_1; 1.
DR Pfam: PF04561; RNA_pol_Rpb2_2; 2.
DR Pfam: PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam: PF00562; RNA_pol_Rpb2_6; 1.
DR Pfam: PF04560; RNA_pol_Rpb2_7; 1.
DR PROSITE: PS01166; RNA_POL_BETA; 1.
KM DNA-directed RNA polymerase; Transcription; Transferase;
KW Complete proteome.
SQ SEQUENCE 1161 AA; 128494 MW; A9C27B928E8191A8 CRC64;

Query Match 27.9%; Score 58; DB 16; Length 1161;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 15; Conservative 3; Mismatches 13; Indels 2; Gaps 2;

QY 4 KTTGMRGHHVEGLAELEQLRRLRLE-HHPQCG 35
DB 227 KALGWTTEQLLEEF-GEYESMRATLEKDHQCG 228

RESULT 14
Q82DQ5 PRELIMINARY; PRT; 1181 AA.
ID 082DQ5;
AC 082DQ5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative RNA polymerase beta subunit.
GN RPOB OR SAV4914.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomyces.
OC NCBI_TaxId=33903;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: APO05040; BAC72626.1; -.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR007121; RNA_pol_B.

```

```

DR InterPro: IPR007644; RNA_pol_Rpb2_1.
DR InterPro: IPR007642; RNA_pol_Rpb2_2.
DR InterPro: IPR007645; RNA_pol_Rpb2_3.
DR InterPro: IPR007120; RNA_pol_Rpb2_6.
DR InterPro: IPR007641; RNA_pol_Rpb2_7.
DR Pfam: PF04563; RNA_pol_Rpb2_1; 1.
DR Pfam: PF04561; RNA_pol_Rpb2_2; 1.
DR Pfam: PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam: PF00562; RNA_pol_Rpb2_6; 1.
DR Pfam: PF04560; RNA_pol_Rpb2_7; 1.
DR PROSITE: PS01166; RNA_POL_BETA; 1.
KW Complete proteome.
SQ SEQUENCE 1181 AA; 130439 MW; AEBF77AC879ED8B CRC64;

Query Match 27.9%; Score 58; DB 16; Length 1181;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 15; Conservative 3; Mismatches 13; Indels 2; Gaps 2;

QY 4 KTTGMRGHHVEGLAELEQLRRLRLE-HHPQCG 35
DB 247 KALGWTTEQLLEEF-GEYESMRATLEKDHQCG 278

RESULT 15
Q7U554 PRELIMINARY; PRT; 229 AA.
ID Q7U554;
AC Q7U554;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SYN1855.
OS Synchococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OC NCBI_TaxId=84588;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641;
RA Palenik B., Brahmsha B., Larimer F.W., Land M., Hauser L.,
RA Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
RA Dufresne A., Patensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synchococcus.";
RL Nature 424:1037-1042(2003).
DR EMBL: BX569694; CAE08370.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 229 AA; 24435 MW; 851E411B69B319C8 CRC64;

Query Match 27.4%; Score 57; DB 16; Length 229;
Best Local Similarity 37.5%; Pred. No. 33;
Matches 12; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 5 TTGMRGHHVEGLAELEQLRRLRLE-HHPQCG 36
DB 81 TSGMREIQAWSGADVALDATORHREGOR 112

```

Search completed: July 15, 2004, 20:35:48
 Job time : 30.5059 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:05 ; Search time 7.15294 Seconds
(without alignments)
276.623 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38
Perfect score: 208
Sequence: 1 MDEKTGMWGSHVVEGLAGELBQRLARLEHNRGQREP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	29.1	177	1 HSLV_RHIL0	Q98C88 rhizobium 1
2	59	28.4	342	1 YDEY_ECOLI	P77672 escherichia
3	57	27.4	474	1 TRPC_CORGL	P06560 coxynebacte
4	56.5	27.2	404	1 KVR3_HUMAN	O43448 homo sapien
5	56.5	27.2	1560	1 SMCK_HUMAN	P41229 homo sapien
6	55.5	26.7	188	1 HSLV_CAUCR	Q9A239 caulobacter
7	54	26.0	376	1 ALG_STRCO	Q9K434 streptomyce
8	53.5	25.7	404	1 KVR3_RAT	O63494 rattus norv
9	53	25.5	348	1 PLX3_SYNY3	P73950 synecocyst
10	53	25.5	589	1 IPEB_ASCSU	P23731 ascaris suu
11	52.5	25.2	184	1 HSLV_BRUME	O8Y311 bruceella me
12	52.5	25.2	184	1 HSLV_BRUSU	O8Y311 bruceella su
13	52	25.0	113	1 XRE_BACSU	P23769 bacillus su
14	52	25.0	356	1 GMB2_CAEEL	Q20636 caenorhabdi
15	52	25.0	499	1 AINX_HUMAN	O16352 homo sapien
16	52	25.0	504	1 AINX_MOUSE	P46660 mus musculu
17	52	25.0	505	1 AINX_RAT	P23565 rattus norv
18	52	25.0	2337	1 TOR2_SCHPO	Q9Y7K2 schizosacch
19	51.5	24.8	181	1 COBU_ECOLI	P46866 escherichia
20	51.5	24.8	273	1 GS80_BACSU	P08078 bacillus su
21	51.5	24.8	401	1 KVR1_MOUSE	O63277 mus musculu
22	51.5	24.8	408	1 KVR1_MUSPU	O88558 mustela put
23	51.5	24.8	419	1 KVR1_HUMAN	O14772 homo sapien
24	51.5	24.8	419	1 KVR1_RABIT	Q9XCT1 oryctolagus
25	51.5	24.8	465	1 FXD3_MOUSE	O61060 mus musculu
26	51	24.5	83	1 RS28_SULSO	Q980G5 sulfolobus
27	51	24.5	261	1 YEVE_YEAST	P40078 saccharomyc
28	51	24.5	358	1 AROB_AERPE	Q9Y319 aeropyrum p
29	51	24.5	429	1 NPX2_MOUSE	O70340 mus musculu
30	51	24.5	431	1 NPX2_HUMAN	P47792 homo sapien
31	51	24.5	432	1 NPX2_RAT	P47792 rattus norv
32	50.5	24.3	149	1 RL13_THEMA	Q9Y135 thermotoga
33	50.5	24.3	176	1 HSLV_RHIME	Q92189 rhizobium m

34	50.5	24.3	1003	1 MBD6_HUMAN	Q96dn6 homo sapien
35	50	24.0	259	1 YAE2_PYPAB	Q9ur44 pyrococcus
36	50	24.0	401	1 KVR3_XENLA	Q9pmt4 xenopus lae
37	50	24.0	473	1 RBL1_RHOCA	O32740 rhodobacter
38	49.5	23.8	366	1 ADH_ALCEU	P14940 alcalligenes
39	49.5	23.8	410	1 AA2A_MOUSE	O60613 mus musculu
40	49.5	23.8	176	1 DEF1_BORPE	Q7wq00 bordetella
41	49	23.6	176	1 DEF2_BORPA	O7w4k0 bordetella
42	49	23.6	241	1 YUO2_ECOLI	P39403 escherichia
43	49	23.6	407	1 PEPT_STRP3	Q8k802 streptococc
44	49	23.6	407	1 PEPT_STRP8	O8p1h9 streptococc
45	49	23.6	407	1 PEPT_STRPY	Q9a0f4 streptococc

ALIGNMENTS

Result ID	HSIV_RHIL0	STANDARD	PRT	177 AA.
AC	Q98C88	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	ATP-dependent protease hslv (EC 3.4.25.-)			
GN	HSIV OR ML5007			
OS	Rhizobium loti (Mesorhizobium loti)			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OX	Phyllobacteriaceae; Mesorhizobium.			
NCBI_TaxID=381;				
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATN-MAF303099;			
RX	MEDLINE=21082950; PubMed=11214968;			
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,			
RA	Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,			
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shingo S., Sugimoto M.,			
RA	Takeuchi C., Yamada M., Tabata S.,			
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium			
RT	Mesorhizobium loti."			
RL	DNA Res. 7:331-338 (2000).			
CC	-1- FUNCTION: Protease subunit of a proteasome-like degradation			
CC	complex (By similarity).			
CC	-1- SUBUNIT: A double ring-shaped homohexameric of hslv is capped on			
CC	each side by a ring-shaped hslu homohexameric (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; AF003005; BAB5153.1; -			
DR	MEROPS; T01.006; -			
DR	HAMAP; MF_00248; -			
DR	InterPro; IPR001353; Peptidase_T1.			
DR	Pfam; PF00227; proteasome; 1.			
KW	Hydrolase; Protease; Threonine protease; Complete proteome.			
FT	ACT_SITE 4 4			
FT	SEQUENCE 177 AA; 18986 MW; 62F3D3DC6ACBSF05 CRC64;			
QY	9 RGHVVEGLAGE-----LEQRLARLEHNP 32			
Db	42 KGGNVLAGPAGATADAFLLERLEATLEQYP 72			

Query Match 29.1%; Score 60.5; DB 1; Length 177;
Best Local Similarity 41.9%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 5; Indels 7; Gaps 1;

```

RESULT 2
YKEY ECOLI
ID YDEY ECOLI STANDARD; PRT; 342 AA.
AC P7672;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter permease protein ydey.
GN YDEY OR B1514.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Bacterichia.
OK NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horinuchi T.;
RT "A 570-bp DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996)
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM EGO/YDEYZ. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
CC THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: Belongs to the binding-protein-dependent transport
CC system permease family. Arah/tpsc subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000249; AAC74587.1; -.
DR EMBL; D90794; BAA15201.1; -.
DR PIR; E64905; E64905.
DR Ecogene; E61807; ydey.
DR InterPro; IPR001851; Bac_inmem_transp.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 14 34
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.

```

```

SQ SEQUENCE 342 AA; 36394 MW; 419E505026ABDE33 CRC64;
Query Match 28.4%; Score 59; DB 1; Length 342;
Best Local Similarity 57.1%; Pred. NO. 2.9;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 8 WRGGHVGAGLGEIQLRL 28
Db 135 WTGGKRWGLPAELKQSLAPL 155

RESULT 3
TRPC CORGL
ID TRPC CORGL STANDARD; PRT; 474 AA.
AC P0650;
DT 01-JAN-1988 (Rel. 06, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypophan biosynthesis protein trpcr [includes: indole-3-glycerol
DE phosphate synthase (EC 4.1.1.48) (IGPS); N-(5'-phospho-
DE ribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI)].
GN TRPC OR CGJ3033.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OK NCBI_Taxid=1718;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87117512; PubMed=3608947;
RA Matsui K., Sano K., Ohtsubo E.;
RT "Complete nucleotide and deduced amino acid sequences of the
RT Brevibacterium lactofermentum tryptophan operon.";
RL Nucleic Acids Res. 14:10113-10114(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: Bifunctional enzyme that catalyzes two sequential steps
CC of tryptophan biosynthetic pathway. The first reaction is
CC catalyzed by the isomerase, coded by the trpF domain; the second
CC reaction is catalyzed by the synthase, coded by the trpC domain.
CC -1- CATALYTIC ACTIVITY: N-(5'-phospho-beta-D-riboseyl)-anthranilate = 1-
CC (2-carboxyphenyl)amino-1-deoxy-D-ribulose 5-phosphate.
CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenyl)amino-1-deoxy-D-ribulose 5-
CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
CC -1- PATHWAY: Tryptophan biosynthesis; third step.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: In the N-terminal section; belongs to the trpC family.
CC -1- SIMILARITY: In the C-terminal section; belongs to the trpF family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04960; BAA28626.1; -.
DR EMBL; AP005283; GAC00427.1; -.
DR PIR; E24723; E24723.
DR HSP; P00909; IPII.
DR HAMAP; MF_00134; fused; 1.
DR HAMAP; MF_00135; fused; 1.
DR InterPro; IPR001468; IGPS.
DR InterPro; IPR001240; PRAI.
DR Pfam; PF00218; IGPS; 1.
DR Pfam; PF00697; PRAI; 1.
DR ProDom; PD001511; IGPS; 1.
DR PROSITE; PS00614; IGPS; 1.

```

KW Trypophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
 KM Decarboxylase; Complete proteome.
 FT DOMAIN 1 262 INDOL-3-GLYCEROL PHOSPHATE SYNTHASE.
 FT 263 474 N-(5'-PHOSPHORIBOSYL)ANTHRANILATE
 FT ISOMERASE.
 FT 88 89 SG -> AA (IN REF. 1).
 FT 110 110 A -> G (IN REF. 1).
 FT 130 131 HA -> RP (IN REF. 1).
 FT 153 153 A -> D (IN REF. 1).
 FT 302 302 L -> S (IN REF. 1).
 FT 343 343 D -> G (IN REF. 1).
 FT 381 383 MISSING (IN REF. 1).
 FT 454 474 AGKADGALKITPATISTFTY -> GWGRCRPAENFRDH
 FT 474 474 LHPLEKV (IN REF. 1).
 SQ SEQUENCE 474 AA; 50477 MW; C347C7016BB97F9A CRC64;
 Query Match 27.4%; Score 57; DB 1; Length 474;
 Best Local Similarity 55.6%; Pred. No. 7.4;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 13 YVEGLAGELEQRLR 30
 DB 13 IVERGRHLEIRARLH 30

RESULT 4
 KVB3_HUMAN STANDARD; PRT; 404 AA.
 AC O43448;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3).
 GN KCNA3B OR KCNA3B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=99074289; PubMed=9857044;
 RA Leicher T., Baehring R., Isbrandt D., Pongs O.;
 RT "Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel A-type potassium channel.";
 RL J. Biol. Chem. 273:35095-35101(1998).
 CC -1- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.5.
 CC -1- SUBUNIT: Forms heteromultimeric complex with alpha subunits.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- TISSUE SPECIFICITY: Brain-specific expression. Most prominent expression in cerebellum. Weaker signals detected in cortex, occipital lobe, frontal lobe and temporal lobe. Not detected in spinal cord, heart, lung, liver, kidney, pancreas, placenta and skeletal muscle.
 CC -1- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).
 CC -1- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; AF016411; AAC92499.1; -.
 CC DR Genew; HGNC:6230; KCNA3B.
 CC MIM; 604111; -.

DR GO; GO:0015459; P:potassium channel regulator activity; TAS.
 DR GO; GO:000613; P:potassium ion transport; TAS.
 DR InterPro; IPR001395; Aldo_ket_red.
 DR InterPro; IPR005402; KCNA3_channel.
 DR InterPro; IPR005399; KCNA3_channel.
 DR InterPro; IPR005983; KCNA3_core.
 DR Pfam; PF00248; aldo_ket_red; 1.
 DR PRINTS; PR01580; KCNA3CHANNEL.
 DR PRINTS; PR01577; KCNA3CHANNEL.
 DR PRODOM; PD000288; Aldo_ket_red; 2.
 DR TIGRfam; TIGR01293; Kv_beta; 1.
 DR TrpAsp; TrpAsp; 1.
 KW Voltage-gated channel.
 KM Ionic channel; Ion transport; Potassium transport;
 SQ SEQUENCE 404 AA; 43530 MW; 08265CC07929A1BA CRC64;
 Query Match 27.2%; Score 56.5; DB 1; Length 404;
 Best Local Similarity 31.1%; Pred. No. 7.2;
 Matches 14; Conservative 5; Mismatches 3; Indels 23; Gaps 2;
 QY 4 KTTGWR-----GG-----HVEGLAGELEQLR 25
 DB 145 KSKGRRSSVYTTIKIFWGQATERGLSRKHLIEGLRSRLQ 189

RESULT 5
 SMCX_HUMAN STANDARD; PRT; 1560 AA.
 AC P41229;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Smcx protein (Xei169 protein).
 GN SMCX OR XE169.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94214434; PubMed=8162017;
 RA Wu J., Ellison J., Salido R., Yen P., Mohandas T., Shapiro L.J.;
 RT "Isolation and characterization of XE169, a novel human gene that escapes X-inactivation.";
 RL Hum. Mol. Genet. 3:153-160(1994).
 RN [2]
 RP SEQUENCE OF 280-344 FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=95038739; PubMed=7951230;
 RA Agulnik A.I., Mitchell M.J., Mattei M.-G., Borsani G., Ayner P.A., Lerner J.L., Bishop C.E.;
 RT "A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human.";
 RL Hum. Mol. Genet. 3:879-884(1994).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P41229-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P41229-2; Sequence=VSP_000315;
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined. Highest levels found in skeletal muscle.
 CC -1- MISCELLANEOUS: Escapes X-inactivation.
 CC -1- SIMILARITY: Contains 1 ARID domain.
 CC -1- SIMILARITY: Contains 1 JMW domain.
 CC -1- SIMILARITY: Contains 1 JMW domain.
 CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

entites requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; 125270; AAA61302.1; -
 CC DR EMBL; 229650; CAA82758.1; -
 CC DR PIR; 154361; 154361.
 DR Genew; HGNC:11114; SMCX.
 DR MIM; 314690; -
 DR InterPro; IPR001606; ARID.
 DR InterPro; IPR003347; TF_JmjC.
 DR InterPro; IPR003349; TF_JmjN.
 DR InterPro; IPR004198; Znf_C5HC2.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF01388; ARID; 1.
 DR Pfam; PF02373; JmjC; 1.
 DR Pfam; PF02375; JmjN; 1.
 DR Pfam; PF0628; PHD; 2.
 DR Pfam; PF02928; Zf_C5HC2; 1.
 DR SMART; SM00501; BRIGHT; 1.
 DR SMART; SM00558; JmjC; 1.
 DR SMART; SM00545; JmjN; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; ZF_PHD_1; 2.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 DR Zinc-finger; Repeat; Alternative splicing.
 FT DOMAIN 13 59 JMJN.
 FT DOMAIN 76 184 ARID.
 FT ZN_FING 326 372 PHD-TYPE 1.
 FT DOMAIN 501 617 JMJN.
 FT ZN_FING 1187 1248 PHD-TYPE 2.
 FT VARBLIC 1370 1372 Missing (in isoform 2).
 FT /FTID=VSP 000315.
 FT C -> Y (IN RFP. 2).
 FT CONFLICT 342 342
 FT SEQUENCE 1560 AA; 175804 MW; 78913F010D08CC61 CRC64;

Query Match 27.2%; Score 56.5; DB 1; Length 1560;
 Best Local Similarity 28.9%; Pred. No. 31;
 Matches 13; Conservative 7; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEKKTGMRG-----GHVVEGLAGELFQRLARLEHHQGQREP 38
 DB 1281 LTERAISWQGRARQALASDVYALIGRLALRQLQAEPRPREPP 1325

RESULT 6
 HSLV CAUCR
 ID HSLV CAUCR STANDARD; PRT; 188 AA.
 AC Q9A239;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP-dependent protease hslv (EC 3.4.25.-).
 GN HSLV OR CC3727.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 CC Caulobacteraceae; Caulobacter.
 RX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 MEDLINE=21173698; PubMed=1259647;
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utecherbeck T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RA "Complete genome sequence of *Caulobacter crescentus*."
 RA Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -I- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (By similarity).
 CC -I- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on

each side by a ring-shaped hslv homohexamer (By similarity).
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -I- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entites requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE006030; AAK25689.1; -
 DR PIR; B87711; B87711.
 DR HSSP; P31059; 1B94.
 DR MEROPS; T01.006; -.
 DR TIGR; CC3727; -.
 DR HAMAP; MF_00248; -; 1.
 DR InterPro; IPR001353; Peptidase_T1.
 DR Pfam; PF00227; proteasome; 1.
 DR Hydrolase; Protease; Threonine protease; Complete proteome.
 FT ACT_SITE 14 14
 FT SEQUENCE 188 AA; 19600 MW; ACCDCIFCB10BD61 CRC64;

Query Match 26.7%; Score 55.5; DB 1; Length 188;
 Best Local Similarity 43.3%; Pred. No. 4.3;
 Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;

QY 10 GGHVVEGLAGE-----LEQRLARLEHHP 32
 DB 52 GGVVAGPAGATADAFLLIERLEAKLEQYP 81

RESULT 7
 ALC_STRCO
 ID ALC_STRCO STANDARD; PRT; 376 AA.
 AC Q9RKU4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative allantoinase (EC 3.5.3.4) (Allantoinase amidinohydrolase).
 GN SC06248 OR SCAN10.13 OR STRA10.13.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Streptomyces; Streptomyces.
 RX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieiser T., Larke L., Larke L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RA "Complete genome sequence of the model actinomycete *Streptomyces*
 RA *coelicolor* A3(2)."
 RL Nature 417:141-147(2002).
 CC -I- CATALYTIC ACTIVITY: Allantoinase + H(2)O = (-)-ureidoglycolate +
 CC urea.
 CC -I- PATHWAY: Degradation of allantoin (purine catabolism); second
 CC step.
 CC -I- SIMILARITY: Belongs to the allantoinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AL939126; CAB60167.1; -
 DR InterPro: IPR005164; Allantoicase.
 DR Pfam: PF03561; Allantoicase; 2.
 KW Hypothetical protein; Hydrolyase; Purine metabolism; Complete proteome.
 SQ SEQUENCE 376 AA; 41207 MW; 2247C0B2300C4A29 CRC64;

Query Match 26.0%; Score 54; DB 1; Length 376;
 Best Local Similarity 44.4%; Pred. No. 14;
 Matches 12; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

QY 10 GGHVEGLAGELEQ--LRARLEHHPOG 34
 DB 171 GGHANGFVAVSAEQRTHRLKQHPDG 197

RESULT 8
 ID KV83 RAT STANDARD; PRT; 404 AA.
 AC 063454;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3) (RCK beta3).
 GN KCNB3 OR CKBETA3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RP TISSUE=Brain cortex;
 RA MEDLINE=96140552; PubMed=8549760;
 RA Heinemann S.H., Rettig J., Wunder F., Pongs O.;
 RT "Molecular and functional characterization of a rat brain Kv beta 3 potassium channel subunit";
 RL FEBS Lett. 377:383-389(1995).
 CC -1- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.4 but not Kv1.1 or Kv1.5.
 CC -1- SUBUNIT: Forms heteromultimeric complex with alpha subunits.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC TISSUE SPECIFICITY: Predominantly expressed in brain. Strongest expression in olfactory bulb and thalamic nuclei. Not detected in heart, spleen, lung, liver, skeletal muscle, kidney and testis.
 CC -1- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).
 CC -1- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

KW Voltage-gated channel.
 SQ SEQUENCE 404 AA; 43689 MW; 7F39B80D15E71A03 CRC64;

Query Match 25.7%; Score 53.5; DB 1; Length 404;
 Best Local Similarity 28.9%; Pred. No. 18;
 Matches 13; Conservative 6; Mismatches 3; Indels 23; Gaps 2;

QY 4 KTTGMR-----GG-----HVEGLAGELEQLR 25
 DB 145 KSKGMRSSSYVTTTKIFWCGAETRGSLRKHITIGLQGLDRDQ 189

RESULT 9
 ID PLSX SYNY3 STANDARD; PRT; 348 AA.
 AC P73950;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Fatty acid/phospholipid synthesis protein p1sx.
 GN PLSX OR SLR1510.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirose M., Sugita M., Sasaki S., Kimura T., Hosouchi T., Matsuo A., Muraki A., Nakazaki N., Nario K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1- FUNCTION: Not known, probably involved in fatty acid or phospholipid synthesis (By similarity).
 CC -1- SIMILARITY: Belongs to the p1sx family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D90911; BAA18017.1; -
 DR PIR: S75456; S75456.
 DR HAMAP: MF_00019; -; 1.
 DR InterPro: IPR003664; FA_synthetis.
 DR Pfam: PF02504; FA_synthetis; 1.
 DR ProDom: PD006974; FA_synthetis; 1.
 DR TIGRFAMs: TIGR00182; p1sx; 1.
 KW Fatty acid biosynthesis; phospholipid biosynthesis; Complete proteome.
 SQ SEQUENCE 348 AA; 37285 MW; 12C07CC3F971A72C CRC64;

Query Match 25.5%; Score 53; DB 1; Length 348;
 Best Local Similarity 34.3%; Pred. No. 18;
 Matches 12; Conservative 9; Mismatches 10; Indels 4; Gaps 2;

QY 2 DEKTTGMRG--GHVVEGLAGELEQLRRLRLEHHPOG 34
 DB 255 EELPRGWRGKLGAIL--LAPNLKRIKQVDAEHRG 287

RESULT 10
 ID IFEB ASCSU STANDARD; PRT; 589 AA.
 AC P23731;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)

```

DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Intermediate filament protein B (IF-B).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoideae).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN (1)
RP SEQUENCE.
RX MEDLINE=90059912; PubMed=2583097;
RA Weber K., Plessmann U., Ulrich W.;
RT "Cytoplasmic intermediate filament proteins of invertebrates are
RT closer to nuclear lamins than are vertebrate intermediate filament
RT proteins; sequence characterization of two muscle proteins of a
RT nematode."
RT EMBL J. 8:3221-3227(1989).
RL
CC -1- SUBUNIT: A and B can form homopolymers.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Giant body muscle cells.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament_1.
DR Pfam: PF00932; IF_tail; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil.
FT DOMAIN 1 84 HEAD.
FT FT 85 433 ROD.
FT FT 434 589 TAIL.
FT DOMAIN 85 116 COIL_1A.
FT FT 117 130 COIL_1A.
FT DOMAIN 131 268 LINKER_1.
FT FT 269 285 COIL_1B.
FT DOMAIN 285 433 LINKER_12.
FT FT 286 433 COIL_2.
SQ SEQUENCE 589 AA; 67694 MW; 5E7D9F6F0AFD56 CRC64;

Query Match 25.5%; Score 53; DB 1; Length 589;
Best Local Similarity 48.0%; Pred. No. 31;
Matches 12; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 3 EKTGKRGHVVEGLAGELEQLRAR 27
Db 182 EKINQWQ-HAIEDAQSELEMLRAR 204

RESULT 11
HSLV_BRUME STANDARD; PRT; 184 AA.
AC Q8YE31;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR BMEI2047.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=16M / ATCC 23456 / Biotype 1;
RC MEDLINE=20020109; PubMed=1156688;
RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Renin G.,
RA Iablonetti L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elser P.H., Hagius S., O'Callaghan D., Letesson T.-J.,
RA Haselkorn R., Kyriades N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RL
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on
each side by a ring-shaped hslv homohexamer (By similarity).

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR DR EMBL; AE009637; AAL3228.1; -.
DR PIR; A13507; A13507.
DR HAMAP; MF_00248; -.
DR InterPro: IPR001353; Peptidase_T1.
DR Pfam; PF00227; proteasome; 1.
KW Hydrolyase; Protease; Threonine protease; Complete proteome.
FT ACT SITE 12 12 BY SIMILARITY.
SQ SEQUENCE 184 AA; 19839 MW; 03EDC7055FDD2EAA7 CRC64;

Query Match 25.2%; Score 52.5; DB 1; Length 184;
Best Local Similarity 41.4%; Pred. No. 10;
Matches 12; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 11 GHVEGLAGE-----LEQLRLRLEHHP 32
Db 51 GNVIQFAGATADATFTLRLRLAKLEQYF 79

RESULT 12
HSLV_BRUSU STANDARD; PRT; 184 AA.
AC Q8FY11;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR BR2080.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=1330 / Biovar 1;
RC MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Frazer C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
RL
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on
CC each side by a ring-shaped hslv homohexamer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR DR EMBL; AE014495; AAN30970.1; -.
DR TIGR; BR2080; -.
DR HAMAP; MF_00248; -. 1.

```

DR InterPro; IPR001353; Peptidase_T1.
 DR Pfam; PF00227; Proteasome_1.
 KM Hydroxylase; Protease; Threonine protease; Complete proteome.
 FT ACT SITE 12 BY SIMILARITY.
 SO SEQUENCE 184 AA, 19809 MW, 1ESB1C68F4A2FEE2 CRC64;
 Query Match 25.2%; Score 52.5; DB 1; Length 184;
 Best Local Similarity 41.4%; Pred. No. 10;
 Matches 12; Conservative 5; Mismatches 5; Indels 7; Gaps 1;
 QY 11 GHVVEGLAGE-----LEQRLARLEHHP 32
 DB 51 GNVYAGFAGATADAFLLERLAKLEQYP 79
 RESULT 13
 XRE_BACSU STANDARD; PRT; 113 AA.
 ID XRE_BACSU
 AC P23789;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE HTH-type transcriptional regulator xre (Putative pbax repressor).
 GN XRE OR BSU12510.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91092505; PubMed=2125016;
 RA Wood H.E., Devine K.M., McConnell D.J.;
 RT "Characterization of a repressor gene (xre) and a
 temperature-sensitive allele from the Bacillus subtilis prophage,
 RT PBX.";
 RL Gene 96:83-86 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168 / SO113;
 RX MEDLINE=94364963; PubMed=8083174;
 RA McDonnell G.E., Wood H., Devine K.M., McConnell D.J.;
 RT "Genetic control of bacterial suicide: regulation of the induction of
 RT PBX in Bacillus subtilis.";
 RL J. Bacteriol. 176:5820-5830 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168;
 RA Krogh S., O'Reilly M., Nolan N., Devine K.M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunze T., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut L.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koehler P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Maesuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parvo V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takenchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).
 CC -1- FUNCTION: PROBABLE REPRESSOR. NECESSARY FOR THE MAINTENANCE OF
 CC THE LYSOGENIC STATE.
 CC -1- MISCELLANEOUS: THE DEFECTIVE PROPHAGE OF BACILLUS SUBTILIS 168,
 CC PBX, IS A CHROMOSOMALLY BASED ELEMENT WHICH ENCODES A
 CC NON-INFECTIONOUS PRAGE-LIKE PARTICLE WITH BACTERICIDAL ACTIVITY.
 CC PBX IS INDUCED BY AGENTS WHICH ELICIT THE SOS RESPONSE.
 CC -1- SIMILARITY: Contains 1 HTH cro/C1-type DNA-binding domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 CC EMBL: M36478; AAA22894.1; -
 CC EMBL: M36477; AAA22893.1; -
 CC EMBL: Z34287; CAA84042.1; -
 CC EMBL: Z70177; CAA94052.1; -
 CC EMBL: Z29110; CAB13108.1; -
 CC PIR: J00810; J00810.
 DR Subtilist; BG10994; xre.
 DR InterPro; IPR001387; HTH_3.
 DR Pfam; PF01381; HTH_3; 1.
 DR SMART; SM00530; HTH_XRE; 1.
 DR PROSITE; PS50943; HTH_CROCI; 1.
 KM Transcription regulation; Repressor; DNA-binding; Complete proteome.
 FT DOMAIN 6 58
 FT DNA_BIND 15 34
 FT VARIANT 4 4
 FT VARIANT 19 19
 FT VARIANT 78 78
 FT VARIANT L -> V (IN TEMPERATURE-SENSITIVE ALLELE).
 SO SEQUENCE 113 AA, 13258 MW, 23AF5C9DE3D2B31A CRC64;
 Query Match 25.0%; Score 52; DB 1; Length 113;
 Best Local Similarity 36.7%; Pred. No. 7.2;
 Matches 11; Conservative 4; Mismatches 15; Indels 0; Gaps 0;
 QY 9 RGHVVEGLAGELEQRLARLEHHPGQREP 38
 DB 10 RGRRTQEEIASHIGVSRRARYSHYENGSRBP 39
 RESULT 14
 GBB2_CAEEL STANDARD; PRT; 356 AA.
 ID GBB2_CAEEL
 AC Q20636;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Guanine nucleotide-binding protein beta subunit 2.
 GN GPB-2 OR F52A8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A., AND INTERACTIONS.
 RX STRAIN=Br18C01 N2;
 RX MEDLINE=21231612; PubMed=11333232;
 RA van der Linden A.M., Simmer F., Cuppen E., Plaetker R.H.A.;
 RT "The G protein beta subunit gpb-2 in Caenorhabditis elegans regulates

```

RT the G(alpha-q)alpha signaling network through interactions with
RT the regulator of G protein signaling proteins egl-10 and eat-16."
RN Genetics 158:221-235(2001).
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Wilkerson J.;
RA Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
CC involved as a modulator or transducer in various transmembrane
CC signaling systems. The beta and gamma chains are required for the
CC GTPase activity, for replacement of GDP by GTP, and for G protein-
CC effector interaction.
CC -1- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
CC gamma. Interacts with goa-1, eat-16, egl-10, egl-30.
CC -1- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF291847; AKK55964.1; -.
DR EMBL; Z71263; CA85824.1; -.
DR PIR; T22478; T22478.
DR HSSP; P04901; ITBG.
DR WormSep; F52A8.2; CE17845.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00082; WD_REPEATS_2; 5.
DR PROSITE; PS50294; WD_REPEAT_REGION; 1.
DR Transducer; Repeat; WD repeat; Multigene family.
DR KW REPEAT 66 WD 1.
DR FT REPEAT 108 WD 2.
DR FT REPEAT 154 WD 3.
DR FT REPEAT 196 WD 4.
DR FT REPEAT 240 WD 5.
DR FT REPEAT 284 WD 6.
DR FT REPEAT 326 WD 7.
DR SEQUENCE 356 AA; 39453 MW; 293AA6DCFA317D832 CRC64;
SQ
Query Match 25.0%; Score 52; DB 1; Length 356;
Best Local Similarity 40.7%; Pred. No. 24;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
QY 3 EKTGWGRGHVVEGLAGELEQLRRLRLE 29
DB 6 QPTTKGSEYVLEQLANEAELRKRLD 32

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=retal brain;
RX MEDLINE=95287809; PubMed=7769995;
RA Chan S.-O., Chiu F.-C.;
RT "Cloning and developmental expression of human 66 kd neurofilament
RT protein."
RT Brain Res. Mol. Brain Res. 29:177-184(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.D., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunnatone P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP FUNCTION: CLASS-IV NEURONAL INTERMEDIATE FILAMENT THAT IS ABLE TO
RP SELF-ASSEMBLE. IT IS INVOLVED IN THE MORPHOGENESIS OF NEURONS. IT
RP MAY FORM AN INDEPENDENT STRUCTURAL NETWORK WITHOUT THE INVOLVEMENT
RP OF OTHER NEUROFILAMENTS OR IT MAY COOPERATE WITH NF-L TO FORM THE
RP FILAMENTOUS BACKBONE TO WHICH NF-M AND NF-H ATTACH TO FORM THE
RP CROSS-BRIDGES.
CC -1- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN ADULT CNS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN BRAIN AS EARLY AS THE 16TH OF
CC GESTATION, AND INCREASED RAPIDLY AND REACHED A STEADY STATE LEVEL
CC BY THE 18TH WEEK OF GESTATION.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S78296; AAB34482.1; -.
DR EMBL; BC006359; AAH06359.1; -.
DR PIR; I52658; I52658.
DR Genew; HGNC:6057; INA.
DR MIM; 605338; -.
DR GO; GO:0005883; C:neurofilament; TAS.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR InterPro; IPR006821; Filament_head.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Filament; 1.
DR Pfam; PF04732; Filament_head; 1.
DR PROSITE; PS00226; IF; 1.
DR KW Intermediate filament; Coiled coil; Neurone.
DR FT DOMAIN 1 87 HEAD.
DR FT DOMAIN 88 408 ROD.
DR FT DOMAIN 409 499 TAIL.
DR FT DOMAIN 88 129 COIL 1A.
DR FT DOMAIN 130 142 LINKER 1.
DR FT DOMAIN 143 238 COIL 1B.
DR FT DOMAIN 239 262 LINKER 2.
DR FT DOMAIN 263 408 COIL 2.

```



```

FT DOMAIN 449 454 POLY-GU.
FT CONFLICT 37 41 GFRSO -> ASVE (IN REF. 1).
FT CONFLICT 67 67 R -> A (IN REF. 1).
FT CONFLICT 92 92 T -> S (IN REF. 1).
FT CONFLICT 128 132 ALRQR -> RCDT (IN REF. 1).
FT CONFLICT 141 141 E -> Q (IN REF. 1).
FT CONFLICT 147 152 LRDIRA -> PRHLP (IN REF. 1).
FT CONFLICT 191 198 GAERALKVA -> RRRLKR (IN REF. 1).
FT CONFLICT 244 244 A -> R (IN REF. 1).
FT CONFLICT 263 263 S -> A (IN REF. 1).
FT CONFLICT 301 301 S -> T (IN REF. 1).
FT CONFLICT 310 311 EE -> DQ (IN REF. 1).
FT CONFLICT 318 318 MISSING (IN REF. 1).
SQ SEQUENCE 499 AA; 55390 MW; 4C972764E9E68D3E CRC64;

```

```

Query Match 25.0%; Score 52; DB 1; Length 499;
Best Local Similarity 47.6%; Pred. No. 35;
Matches 10; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 15 EGIAGELBOLRLRLEHHPOGO 35
   :||| |::||| | :| :
Db 169 DGIABEVQRLRARCCEBSRGR 189

```

Search completed: July 15, 2004, 20:34:29
 Job time : 8.15294 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:31:20 ; Search time 10.2824 Seconds
(without alignments)
355.490 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208

Sequence: 1 MDEKTTGMRGHVVEGLAGELRRLRLHHPQCGREP 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	28.4	342	2	B64905
2	59	28.4	342	2	A90894
3	59	28.4	342	2	H85723
4	57	27.4	474	1	B24723
5	57	27.4	663	2	D83545
6	56.5	27.2	1560	2	IS4361
7	55.5	26.7	188	2	E87711
8	54	26.0	321	2	H87683
9	54	26.0	1156	2	B69444
10	53.5	25.7	404	2	S68409
11	53.5	25.7	554	2	E87375
12	53	25.5	261	2	C75568
13	53	25.5	348	1	S75456
14	53	25.5	496	2	H70658
15	53	25.5	589	2	S06954
16	53	25.5	947	2	E86362
17	52.5	25.2	184	2	AI3507
18	52.5	25.2	392	2	P84650
19	52.5	25.2	744	2	A82822
20	52	25.0	113	2	JU0810
21	52	25.0	356	2	T22478
22	52	25.0	376	2	T34734
23	52	25.0	494	2	I52658
24	52	25.0	504	2	I53868
25	52	25.0	505	2	A41023
26	52	25.0	636	2	F47236
27	52	25.0	2337	2	T40577
28	51.5	24.8	134	2	AH3177
29	51.5	24.8	181	2	H64963

30	51.5	24.8	181	2	D90977
31	51.5	24.8	181	2	B85824
32	51.5	24.8	273	2	G69770
33	51.5	24.8	308	2	P95903
34	51.5	24.8	401	2	S66503
35	51.5	24.8	408	2	I59393
36	51.5	24.8	419	2	I55463
37	51.5	24.8	1156	2	A29838
38	51	24.5	98	2	F90164
39	51	24.5	169	2	F83372
40	51	24.5	228	2	C90250
41	51	24.5	261	2	S43218
42	51	24.5	351	2	D88065
43	51	24.5	358	2	C72643
44	51	24.5	429	2	F84015
45	51	24.5	432	2	T10894

ALIGNMENTS

RESULT 1
B64905
probable sugar transport permease protein b1514 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: B64905
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9276503
A:Accession: B64905
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-342 <BLAT>
A:Cross-references: GB:AE000449; GB:U00096; NID:G1787790; PIDN:AC74587.1; PID:G1787793,
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: 1-arabinose transport system permease araH
C:Keywords: transmembrane protein
F:11-27/Domain: transmembrane #status predicted <TM1>
F:44-60/Domain: transmembrane #status predicted <TM2>
F:65-81/Domain: transmembrane #status predicted <TM3>
F:88-104/Domain: transmembrane #status predicted <TM4>
F:107-123/Domain: transmembrane #status predicted <TM5>
F:155-171/Domain: transmembrane #status predicted <TM6>
F:208-224/Domain: transmembrane #status predicted <TM7>
F:247-263/Domain: transmembrane #status predicted <TM8>
F:266-282/Domain: transmembrane #status predicted <TM9>
F:291-307/Domain: transmembrane #status predicted <TM10>

Query Match 28.4%; Score 59; DB 2; Length 342;
Best Local Similarity 57.1%; Pred. No. 5.4;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 8 MRGHVVEGLAGELRRLRLHHPQCGREP 28
Db 135 WTGKWTBGLPAELKQLSLPL 155

RESULT 2
A90894
probable transpore system permease protein Ece2121 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: A90894
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasegawa, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A90894
A:Status: preliminary

A: Molecule type: DNA
A: Residues: 1-342 <HAY>
A: Cross-References: GB:BA000007, PIDN:EBB35544.1, PID:gl3361587, GSPDB:GN00154
A: Experimental source: strain O157:H7, substrain RIMD 0509952
C: Genetics:
C: Gene: ECs2121
C: Superfamily: 1-arabinose transport system permease araH

Query Match	28.4%	Score 59;	DB 2;	Length 342;
Best Local Similarity	57.1%	Pred. No. 5.4;		
Matches 12;	Conservative 2;	Mismatches 7;	Indels	

QY 8 WRGGHVEGLAGELEQLRRL 28
 135 WTGGKMIIEGLPAELKQLSAPL 155
 Db

RESULT 3
H85723
probable transport system permease protein ydaY [imported] - Escherichia coli (strain O1

C/Species: Escherichia coli
C/date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: H85723

R. Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamounsis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: H85723

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-342 <STO>
```

A:Cross-references: GB:AB005174; NID:G1251515; PIDN:AAG56252.1; GSPDB:IGN00145; UMG9:Z21
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:

A/Gene: ydeY
C/Superfamily: 1-arabinose transport system permease arah

Query Match	28.4%	Score 59;	DB 2;	Length 342;
Best Local Similarity	57.1%;	Pred. No. 5.4;		
Matches 12; Conservative	2;	Mismatches	7;	Indels 0;
				Gaps 0;

```
QY      8 WRGCHVVEGLAGELEQLRRL 28
      ||| : ||| ||: ||| |||
Db     135 WTGCKWIEGLPAELKQLSAPL 15
```

RESULT 4
E24723

tryp protein - *Corynebacterium glutamicum*
C/Species: *Corynebacterium glutamicum*
C/Date: 10-Sen-1999 #sequence revision 10-Sen-1999 #ext chance 10-Sen-1999

C/Accession: E24723
R;Matsui, K.; Sano, K.; Ohtsubo, E.
Nucleic Acids Res. 14: 10113-10114. 1986

A:Title: Complete nucleotide and deduced amino acid sequences of the *Brevibacterium lact*
A:Reference number: A93606; MUID:87117512; PMID:3808947
A:Contents: B lactofermentum

A;Accession: E24723
 A;Molecule type: DNA
 A;Residues: 1=474 (MAT)

A; Cross-references: GB:X04960; NID:g39591; PIDN:CAA28626.1; PID:g39596
C; Genetics:
A; Gene: tpmC

C/Superfamily: tRPC-ctRF bifunctional enzyme; tRPC homology; ctRF homology
F/9-258/Domain: tRPC homology <TRC>
F/354-460/Domain: tRF homology <TRF>

Query Match	27.4%	Score 57	DB 1	Length 474
Best Local Similarity	55.6%	Pred. No. 14		
Matches 10	Conservative 4	Mismatches 4	Indels 0	Gaps 0

QY 13 VEGELGELEQRLARLEH 30

```
Db      13 IVEGRGHLEERARIAH 30
```

RESULT 5

D83545
probable helicase PA0799 (imported) - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*

CiDate: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83545
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizouchi, S.D.; Warren, P.; Hic

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: D83545

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-663 <ST

A:Cross-references: GB:AE004515, GB:AE004091, NID:g93946687, PIDN:AAG04188.1, GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:

A;Gene: PA0799

Query Match	27.4%	Score 57;	DB 2;	Length 663;
Best Local Similarity	36.1%	Pred. No. 20;		
Matches 13; Conservative	3;	Mismatches 20;	Indels 0;	Gaps 0;

```
Qy      1 MDEKTTGMRGGHVEGLAGELEQLRLRLEHHPRQGR 36
      :::::|::|
Db      154 LDELPLAEGGGDNIKDFARRLRSQARPATPRPQGR 189
```

RESULT 6
I54361

SMCX protein - human
N;Alternate names: escapes X-chromosome inactivation
C;Species: Homo sapiens (man)

CjDate: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Mar-2000
CjAccession: I54361; I38235
RjNu, J.; Ellison, J.; Salido, E.; Yen, P.; Mohandas, T.; Shapiro, L.J.

Hum. Mol. Genet. 3, 153-160, 1994
A; Title: Isolation and characterization of XE169, a novel human gene th
A; Reference number: 154361; MUID: 94214434; PMID: 8162017

A:Accession: I54361
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A/Residues: 1-1560 <WTJ>
A/Cross-references: GB:L25270; NID:g457136; PID:g457137
R:Aquilik, A.I.; Mitchell, M.J.; Mattei, M.G.; Borsani,

Hum. Mol. Genet. 3, 879-884, 1994
 A:Title: A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation
 A:Reference number: 138235; MUID:95038739; PMID:7951230

A1:Accession: I38235
A1:Status: preliminary;
A1:Molecule type: mRNA
translated from GB/EMBL/DBJ

A:Residues: 280-341, 'Y', 343-344 <AGU>
A:Cross-references: EMBL:Z29650, NID:g472994, PIDN:CAA82758.1, PID:g472995
C:Genetics:

A1:Gene: GDB:SMCX; XE169; DXS1272E
A1:Cross-references: GDB:9865642; OMIM:314690
A1:Map position: Xp11.22-Xp11.21

Query Match 27.2% Score 56.5 DB 2: Length 1560:
C:Superfamily: human retinoblastoma binding protein 2

Best Local Similarity	28.9%;	Pred. No. 58;
Matches	13;	Conservative
	7;	Mismatches
	18;	Indels
	7;	Gaps
	1.	

```
QY      1 MDEKTTGWRG-----GHVVEGLAGELEQLRLARLEHHHPQGRPE 38
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1281 LTERAISWGGRARQALASEVDYALLGRLAELRQRLQAEPRPEEP 1325
```

RESULT 7
E87711
heat shock protein HslV [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: E87711
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: E87711
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-188 <STO>
A/Cross-references: GB:AE005673; NID:G13425497; PIDN:AAK25689.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC3727

Query Match 26.7%; Score 55.5; DB 2; Length 188;
Best Local Similarity 43.3%; Pred. No. 8.2;
Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;

QY 10 GGHVVEGLAGE-----LEQLRARELHHHP 32
DB 52 GGRVAVGFGATATDATTLERLAKLEQYF 81

RESULT 8
H87683
transcription regulator, Arac family [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: H87683
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: H87683
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-321 <STO>
A/Cross-references: GB:AE005673; NID:G13425234; PIDN:AAK25468.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC3506

Query Match 26.0%; Score 54; DB 2; Length 321;
Best Local Similarity 30.3%; Pred. No. 23;
Matches 10; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

QY 2 DEKTGMRGHVVVEGLAGELEQLRARELHHHP 34
DB 205 DEDRVMAAAQIRGAGGTIDRLAKRIDAPRG 237

RESULT 9
E69444
chromosome segregation protein (smc1) homolog - Archaeoglobus fulgidus
C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Jun-2000
C/Accession: E69444
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A/Authors: Uetzelback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaime, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: E69444
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1156 <KLE>
A/Cross-references: GB:AE000995; GB:AE000782; NID:G2689318; PIDN:AA896690.1; PID:G26490C
C/Superfamily: chromosome segregation protein SMC1

Query Match 26.0%; Score 54; DB 2; Length 1156;
Best Local Similarity 44.4%; Pred. No. 89;
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 11 GHVVEGLAGELEQLRARELHHHPGORE 37
DB 349 GEIVDEIAAKWELRQRELVYDKHRE 375

RESULT 10
S68409
potassium channel Kv beta-3 (chain - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: S68409; S72562
R/Heinemann, S.H.; Rettig, J.; Wunder, F.; Pongs, O.
FEBS Lett. 377, 383-389, 1995
A/Title: Molecular and functional characterization of a rat brain K(v)-beta3 potassium c
A/Reference number: S68409; MUID:96140552; PMID:8549760
A/Accession: S68409
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-404 <HEI>
A/Cross-references: EMBL:X76723
A/Experimental source: cortex
R/Rettig, J.
submitted to the EMBL Data Library, December 1993
A/Reference number: S72562
A/Accession: S72562
A/Molecule type: mRNA
A/Residues: 1-245, 'F', 247-324, 'K', 326-404 <RET>
A/Cross-references: EMBL:X76723; NID:G1246833; PIDN:CAAS4141.1; PID:G1246834
A/Experimental source: cortex
C/Genetics:
A/Gene: RCKDelta3
C/Superfamily: fission yeast pyridoxine 4-dehydrogenase
C/Keywords: voltage-gated ion channel

Query Match 25.0%; Score 53.5; DB 2; Length 404;
Best Local Similarity 28.9%; Pred. No. 34;
Matches 13; Conservative 6; Mismatches 3; Indels 23; Gaps 2;

QY 4 KTTGMR-----GG-----HVVVEGLAGELEQLR 25
DB 145 KSKGRSSSVYITTKIFWGGQAEYERGLSRKHITGLQSLDRIQ 189

RESULT 11
E87375
conserved hypothetical protein CC1017 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: E87375
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: E87375
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-554 <STO>
A/Cross-references: GB:AE005673; NID:G13422309; PIDN:AAK23001.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC1017

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:32:20 ; Search time 12.9647 Seconds
(without alignments)
151.318 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38
Perfect score: 208
Sequence: 1 MDEKTTGWRGVHVEGLAGLEQLRLARLEHHHPCQGREP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/1aa/5A COMB pep:.*
2: /cgn2_6/prodata/2/1aa/5B COMB pep:.*
3: /cgn2_6/prodata/2/1aa/5A COMB pep:.*
4: /cgn2_6/prodata/2/1aa/5B COMB pep:.*
5: /cgn2_6/prodata/2/1aa/5A COMB pep:.*
6: /cgn2_6/prodata/2/1aa/5B COMB pep:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	30.3	543	3	US-09-413-814-91 Sequence 91, Appl
2	63	30.3	544	3	US-09-413-814-81 Sequence 81, Appl
3	59.5	28.6	543	4	US-09-252-991A-29830 Sequence 29830, A
4	58.5	28.1	197	4	US-09-252-991A-24825 Sequence 24825, A
5	58	27.9	113	2	US-08-248-839C-6 Sequence 6, Appl
6	58	27.9	628	4	US-09-252-991A-22131 Sequence 22131, A
7	57.5	27.6	239	4	US-09-134-000C-5025 Sequence 5025, Ap
8	57	27.4	356	4	US-09-252-991A-19549 Sequence 19549, A
9	57	27.4	904	4	US-09-252-991A-29119 Sequence 29119, A
10	55.5	26.7	1388	4	US-09-572-191-2 Sequence 2, Appl
11	55.5	26.7	1388	4	US-09-723-262-2 Sequence 2, Appl
12	55.5	26.7	1388	4	US-09-723-219-2 Sequence 2, Appl
13	54	26.0	185	4	US-09-252-991A-19647 Sequence 19647, A
14	53.5	25.7	266	4	US-09-252-991A-24870 Sequence 24870, A
15	53.5	25.7	622	4	US-09-252-991A-19069 Sequence 19069, A
16	53	25.5	152	4	US-09-252-991A-18469 Sequence 18469, A
17	53	25.5	485	4	US-08-311-731A-3 Sequence 3, Appl
18	52.5	25.2	69	4	US-09-328-352-4132 Sequence 4132, Ap
19	52.5	25.2	220	4	US-09-252-991A-24796 Sequence 24796, A
20	52.5	25.2	279	4	US-09-252-991A-19165 Sequence 19165, A
21	52.5	25.2	283	4	US-09-252-991A-20071 Sequence 20071, A
22	52.5	25.2	459	4	US-09-134-000C-6639 Sequence 6639, Ap
23	52.5	25.2	524	4	US-09-252-991A-19671 Sequence 19671, A
24	52.5	25.2	631	4	US-09-252-991A-26444 Sequence 26444, A
25	52	25.0	113	4	US-08-248-839C-2 Sequence 2, Appl
26	52	25.0	113	2	US-08-248-839C-8 Sequence 8, Appl
27	52	25.0	152	4	US-09-252-991A-31405 Sequence 31405, A

28	52	25.0	261	4	US-09-252-991A-28545 Sequence 28545, A
29	52	25.0	423	2	US-08-715-554-2 Sequence 2, Appl
30	52	25.0	423	2	US-08-583-118-2 Sequence 2, Appl
31	52	25.0	499	1	US-09-252-991A-27221 Sequence 27221, A
32	52	25.0	741	4	US-09-252-991A-27062 Sequence 27062, A
33	51.5	24.8	329	2	US-08-606-143-1 Sequence 31496, A
34	51.5	24.8	329	2	US-08-606-143-1 Sequence 3, Appl
35	51.5	24.8	329	2	US-08-606-143-3 Sequence 3, Appl
36	51.5	24.8	478	4	US-09-328-352-5059 Sequence 5059, Ap
37	51	24.5	166	4	US-09-252-991A-24661 Sequence 24661, A
38	51	24.5	179	4	US-09-252-991A-30679 Sequence 30679, A
39	51	24.5	293	4	US-09-252-991A-27745 Sequence 27745, A
40	51	24.5	248	4	US-09-252-991A-28815 Sequence 28815, A
41	51	24.5	324	4	US-09-489-039A-10737 Sequence 10737, A
42	51	24.5	427	4	US-09-252-991A-20199 Sequence 20199, A
43	51	24.5	428	4	US-08-631-607-7 Sequence 7, Appl
44	51	24.5	430	1	US-09-098-358B-7 Sequence 7, Appl
45	51	24.5	430	1	US-09-098-358B-7 Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-09-413-814-91
; Sequence 91, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bioeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofer, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE OF INVENTION: heteropolypeptide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413, 814
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-91

Query Match
Best Local Similarity 30.13% Score 63; DB 3; Length 543;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY 8 WRGVHVEG-----LAGELEQLRLARLEHHHPCQGR 36
Db 235 WLSGEVLSEGLKLVYTKLSGALRRARVPVDEHPAGR 271

RESULT 2
US-09-413-814-81
; Sequence 81, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bioeker, Helmut
; APPLICANT: Brandt, Petra
```

APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 81
LENGTH: 544
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-81

Query Match 30.3%; Score 63; DB 3; Length 544;
Best Local Similarity 35.1%; Pred. No. 0.97;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY 8 WRGHHVEG-----LAGELEQLRLRHHHPQGR 36
DB 236 WLSEVLEGBELRYWTKLSGALRRARVPDHPGR 272

RESULT 3
US-09-252-991A-29830
Sequence 29830, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29830
LENGTH: 543
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29830

Query Match 28.6%; Score 59.5; DB 4; Length 543;
Best Local Similarity 42.9%; Pred. No. 3;
Matches 15; Conservative 3; Mismatches 14; Indels 3; Gaps 2;

QY 7 GWRGHHVEGLAG--ELEQLRLRHHHPQGR 38
DB 183 GQAGHAFGLADLVLDLGVAGRRHPRPDRMP 217

RESULT 4
US-09-252-991A-24825
Sequence 24825, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24825
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24825

Query Match 28.1%; Score 58.5; DB 4; Length 197;
Best Local Similarity 44.8%; Pred. No. 1.3;
Matches 13; Conservative 2; Mismatches 13; Indels 1; Gaps 1;

QY 8 WRGHHVEGLAGELEQLRLRHHHPQGR 35
DB 112 WRGHAEPDGPCHRRRRLPLRRHHHPGR 140

RESULT 5
US-08-248-839C-6
Sequence 6, Application US/08248839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McConnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 113 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-248-839C-6

Query Match 27.9%; Score 58; DB 2; Length 113;
Best Local Similarity 40.0%; Pred. No. 0.81;
Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 9 RGHVVEGLAGLEQLRLRHHHPQGR 38
DB 10 RGRQERIAHGIGVRSARYSHYENGRESE 39

RESULT 6
US-09-252-991A-22131
Sequence 22131, Application US/09252991A
Patent No. 6551795


```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22131
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22131

Query Match      27.4%; Score 58; DB 4; Length 628;
Best Local Similarity 44.4%; Pred. No. 5.8;
Matches 16; Conservative 1; Mismatches 11; Indels 8; Gaps 1;

QY      9 RGGHVE-----GLAGEQLRLARLEHHPOGQR 36
Db      126 RGGHAEVVGGLAVGGGLAVGLGAVPGRGADHPVGPR 161

RESULT 7
US-09-134-000C-5025
; Sequence 5025, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5025
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5025

Query Match      27.6%; Score 57.5; DB 4; Length 239;
Best Local Similarity 52.2%; Pred. No. 2.3;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY      8 WRGHHVEGLAGEQLRLARLEH 30
Db      34 W-GEVITGIMASMEQERAKEH 55

RESULT 8
US-09-252-991A-19549
; Sequence 19549, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 19549
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19549

Query Match      27.4%; Score 57; DB 4; Length 356;
Best Local Similarity 45.5%; Pred. No. 4.2;
Matches 15; Conservative 3; Mismatches 9; Indels 6; Gaps 2;

QY      7 GWRGHHVEGLAGEQLRLARLEHHPO--GQRE 37
Db      183 GRRGGHAGGEPQ-----RATGQHPHVAVGQRQ 211

RESULT 9
US-09-252-991A-29119
; Sequence 29119, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29119
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29119

Query Match      27.4%; Score 57; DB 4; Length 904;
Best Local Similarity 36.1%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY      1 MDEKTTGWRGHHVEGLAGEQLRLARLEHHPOGQR 36
Db      395 LDELPLAWGCGDNLDFARLRLSFQARPATPPQGR 430

RESULT 10
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466 motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2

Query Match      26.7%; Score 55.5; DB 4; Length 1388;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 12; Conservative 6; Mismatches 7; Indels 7; Gaps 1;

QY      9 RGGHVE-----GLAGEQLRLARLEHHPO 33

```

Db 478 RCGFLPEODRLSELRLNEIQTLEQIEHHPR 509

RESULT 11

US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2

Query Match 26.7%; Score 55.5; DB 4; Length 1388;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 12; Conservative 6; Mismatches 7; Indels 7; Gaps 1;

Qy 9 RCGHVE-----GLAGLEQLRLARLEHHPR 33

Db 478 RCGFLPEODRLSELRLNEIQTLEQIEHHPR 509

RESULT 12

US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-219-2

Query Match 26.7%; Score 55.5; DB 4; Length 1388;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 12; Conservative 6; Mismatches 7; Indels 7; Gaps 1;

Qy 9 RCGHVE-----GLAGLEQLRLARLEHHPR 33

Db 478 RCGFLPEODRLSELRLNEIQTLEQIEHHPR 509

RESULT 13

US-09-252-991A-19647
; Sequence 19647, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19647
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19647

Query Match 26.0%; Score 54; DB 4; Length 185;
Best Local Similarity 40.5%; Pred. No. 5.2;
Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;

Qy 2 DEKTGMRGHVEGLAGLEQLRLARLEHH--PQQ 35

Db 98 DEMERGECISVPGLRGAVSRHR-RIRYQGLDPOGQ 133

RESULT 14

US-09-252-991A-24870
; Sequence 24870, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24870
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (211)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-24870

Query Match 25.7%; Score 53.5; DB 4; Length 266;
Best Local Similarity 34.2%; Pred. No. 9.3;
Matches 13; Conservative 4; Mismatches 6; Indels 15; Gaps 1;

Qy 15 EGLAGLEQLRLARLEHHPR-----OGORE 37

Db 112 QGVEPEQLKQAVDHOHQALLPCREHQAQLVGGQHE 149

RESULT 15

US-09-252-991A-19069
; Sequence 19069, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ. ID NOS: 33142
 ; SEQ. ID NO 19069
 ; LENGTH: 692
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-19069

Query Match 25.7%; Score 53.5; DB 4; Length 692;
 Best Local Similarity 42.9%; Pred. No. 28;
 Matches 12; Conservative 4; Mismatches 9; Indels 3; Gaps 1;

QY 6 TGRGGHVVEGLAGELQLEARLEHHPQ 33
 DB 287 TGRHVGSEYIEGVSIGRLQRL--AERHPR 311

Search completed: July 15, 2004, 20:37:00
 Job time : 12.9647 secs

This Page Blank (uspto)

This Page Blank (uspto)